BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

EDITED BY

EGON S. PEARSON

IN CONSULTATION WITH

HARALD CRAMÉR
R. C. GEARY
MAJOR GREENWOOD
J. B. S. HALDANE
G. M. MORANT
JOHN WISHART

VOLUME XXXIV

1947

ISSUED BY THE BIOMETRIKA OFFICE
UNIVERSITY COLLEGE, LONDON
AND PRINTED AT THE
UNIVERSITY PRESS, CAMBRIDGE

PRINTED IN GREAT BRITAIN

CONTENTS OF VOLUME XXXIV

Memoirs

Contents

Miscellanea

VOLUME XXXIV, PARTS I AND II JANUARY 1947

THE VARIANCE OF THE OVERLAP OF GEOMETRICAL FIGURES WITH REFERENCE TO A BOMBING PROBLEM

BY F. GARWOOD, PH.D.

1. INTRODUCTION

The present paper deals with a particular problem arising in the mathematical study of bombing. Briefly, the general problem is that of predicting the over-all effects of a bombing attack carried out under given conditions against a given target, and the mathematical treatment involves various simplifying assumptions concerning these conditions.

In the type of problem considered here, attention is centred on the total plan area of damage caused to a single building by bombs falling independently and at random over a larger area containing the building. It is assumed that each bomb damages all that part of the building contained within a circle of fixed size centred at the bomb (a square damage area is also considered), while the building has a simple plan outline, such as a rectangle or a circle. The area of damage of two or more adjacent bombs is merely the area covered by the circles. The theoretical problems dealt with are those of estimating the variance of the amount of damaged area (the estimation of the mean or expected damage presents little difficulty). It would be more satisfactory to obtain the complete frequency distributions, but this has so far not been achieved, nor has it been possible to obtain explicit formulae for the 3rd and 4th moments.

As there may be applications of the problems to fields completely different from those of bombing studies, and as they are problems which involve essentially the concepts of geometry and of probability, it is convenient to express them entirely in these terms.

We thus have problems of the following type. A number of circles are placed at random on a plane so that each one has some or all of its area inside a fixed square. What are the mean and variance of the area of the square covered by the circles? The fundamentals of this type of problem have been studied by Robbins (1944), and Bronowski & Neyman have dealt with another particular case.* Robbins's results enable us to deal with geometrical figures other than circles and squares, and also to deal with cases where the number, position and orientation of the 'covering' figures follow probability laws other than the simple ones implied in the above example.

2. ROBBINS'S THEOREM

In leading up to his theorem, Robbins uses the concept of a random measurable subset X of n-dimensional Euclidean space E_n. He defines the function $g(x, X)$ for every point x of E_n and for every X as equal to 1 for $x \in X$ and zero elsewhere. This theorem is then as follows:

Let X be a random Lebesgue measurable subset of E_n, with measure $\mu(X)$. For any point x of E_n let $p(x) = Pr(x \in X)$. Then, assuming that the function $g(x, X)$ is a measurable function of the pair (x, X), the expected value of the measure of X will be given by the Lebesgue integral of the function $p(x)$ over E_n.

* *Note by Editor.* This paper was received for publication in September 1945; Dr Garwood has asked me to add the following note in proof. "The author had the privilege of seeing the work of Bronowski & Neyman in proof. This paper was then submitted, after which their work was published (1945) together with a second article by Robbins (1945), who has solved, among others, some of the problems dealt with in this paper, as acknowledged in later footnotes."

Robbins generalizes this result to obtain the mth moment of the measure of X; this is the integral of the function $p(x_1, x_2, \ldots, x_m)$ over E_{mn}, where

$$p(x_1, x_2, \ldots, x_m) = Pr(x_1 \in X \text{ and } x_2 \in X \ldots \text{ and } x_m \in X). \tag{1}$$

It is useful to give a simple non-rigorous proof of this result. Suppose the space E_n to be divided into an enumerably infinite set of small elements $\omega_1, \omega_2, \ldots$. If we assume that any particular subset X can be made up of a selection of the ω's, then

$$\mu(X) = \lambda_1\omega_1 + \lambda_2\omega_2 + \ldots,$$

where ω_i is here used also as the measure of the element ω_i, and where the λ's, appropriate to this particular X, are 0 or 1. Hence

$$\{\mu(X)\}^m = \sum_p \sum_q \ldots \lambda_p \lambda_q \ldots \omega_p \omega_q \ldots,$$

where each summation is over the whole of E_n, and there are m such summations. Thus

$$\exp\{\mu(X)\}^m = \sum_p \sum_q \ldots \omega_p \omega_q \ldots \exp(\lambda_p \lambda_q \ldots).$$

But the expectation of $\lambda_p \lambda_q \ldots$ is the probability that the elements $\omega_p, \omega_q, \ldots$ are in X, and on proceeding to the limit the desired result is obtained.

The verification of this result in the case, say, of the 2nd moment of a linearly distributed variate, is instructive. Thus suppose x is a variate with a probability function $F(x)$, i.e. the probability of obtaining a value $\leqslant x$ is given by the measurable function $F(x)$, where

$$F(-\infty) = 0 \quad \text{and} \quad F(\infty) = 1.$$

Define X as the interval from 0 to x; then the expectation of the square of the measure of X is the 2nd moment of x. To use Robbins's theorem we used the probability $p(x_1, x_2)$ that a given pair of values x_1 and x_2 both lie in the interval $0, x$ chosen at random. Using co-ordinate axes Ox_1, Ox_2, this probability is zero in the 2nd and 4th quadrants, since O, x cannot contain two points x_1 and x_2 of opposite signs. In the region A (see Fig. 1), where $x_1 > x_2 > 0$, the two values x_1 and x_2 are both in O, x if $x > x_1$, and the probability of this is $1 - F(x_1)$. Thus in A,

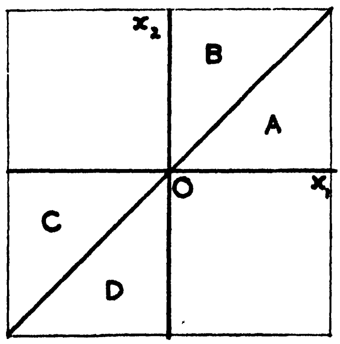


Fig. 1.

$$p(x_1, x_2) = 1 - F(x_1).$$

Similarly in B $\quad p(x_1, x_2) = 1 - F(x_2)$,

while in C $\quad p(x_1, x_2) = F(x_1)$

and in D $\quad p(x_1, x_2) = F(x_2)$.

The integral of $p(x_1, x_2)$ over A is seen to be

$$\int_0^\infty x_1[1 - F(x_1)]\, dx_1,$$

while the total integral of $p(x_1, x_2)$ over the whole plane is

$$2\int_0^\infty x[1 - F(x)]\, dx - 2\int_{-\infty}^0 xF(x)\, dx.$$

A single integration by parts then leads to

$$\int_{-\infty}^\infty x^2\, dF(x),$$

which is the 2nd moment of x, as required.

3. APPLICATION OF ROBBINS'S THEOREM TO OVERLAP PROBLEMS

We shall be concerned with cases in E_2 where the subset X is the part of a fixed area A in the plane which is covered by a number of areas C dropped independently and at random on the plane. We suppose A to be the interior of a simple closed curve, while each C is the interior of another curve. The area C has a reference point Q (conveniently called its centre) and a reference line, and it is assumed that there is a frequency distribution $\phi(x,y,\theta)$ of the position (x,y) of Q and of the inclination θ of the reference line to a fixed direction in E_2. $\phi(x,y,\theta)$ can be assumed to be zero outside an area T, i.e. the points Q are distributed inside T. (In the applications the angle θ will be constant and the areas C will be equally likely to fall anywhere over T, so that we can write $\phi(x,y,\theta) = 1/T$.) Another chance variable is k, the number of areas C; its distribution can be defined by the series $p_0, p_1, p_2, \ldots, p_k, \ldots$ (which are the probabilities of $0, 1, 2, \ldots, k, \ldots$ C's falling on T), or by the probability generating function $G(u)$, where

$$G(u) \equiv p_0 + p_1 u + p_2 u^2 + \ldots + p_k u^k + \ldots. \tag{2}$$

Finally, it is more convenient to consider the moments of the area $Y = A - X$, i.e. the area of A (we can use the symbols Y, etc. for either the sets or their areas) not covered by the C's. Evidently the variances of X and Y are equal.

To obtain the 1st moment of Y, we need first the probability $p(x_1,y_1)$ that a point (x_1,y_1) of A will belong to Y, i.e. of (x_1,y_1) not being covered by a C. Now (x_1,y_1) will not be covered by a particular C falling at an inclination θ if the centre $Q(x,y)$ falls outside an area $\bar{C}(x_1,y_1,\theta)$ obtained by centring the C at (x_1,y_1) and rotating it through 180°. If the part of T exterior to this area is called $T-\bar{C}(x_1,y_1,\theta)$, and if we allow all inclinations, the probability of this occurring is

$$q(x_1,y_1) = \int_0^{2\pi} \int \int_{T-\bar{C}(x_1,y_1,\theta)} \phi(x,y,\theta)\,dx\,dy\,d\theta. \tag{3}$$

If k C's are dropped independently, the probability is given by $q^k(x_1,y_1)$, so that the total probability of (x_1,y_1) belonging to Y is

$$p(x_1,y_1) = \sum_{k=0}^{\infty} p_k q^k(x_1,y_1) = G\{q(x_1,y_1)\}. \tag{4}$$

The 1st moment of Y is thus, in the case of k C's,

$$\mu_1'(Y) = \int\int_A q^k(x_1,y_1)\,dx_1\,dy_1, \tag{5}$$

and in the general case

$$\mu_1'(Y) = \int\int_A G\{q(x_1,y_1)\}\,dx_1\,dy_1. \tag{6}$$

The 2nd moment is obtained by a similar process; we require the probability $p(x_1,y_1,x_2,y_2)$ that neither of two points (x_1,y_1) and (x_2,y_2) is covered by a C. Corresponding to these two points and an inclination θ the permissible region in which each centre Q can fall is

$$T - \bar{C}(x_1,y_1,\theta) - \bar{C}(x_2,y_2,\theta) \equiv T - \bar{C}_1 - \bar{C}_2.$$

Thus for one C the probability is

$$q(x_1,y_1,x_2,y_2) = \int_0^{2\pi} \int \int_{T-\bar{C}_1-\bar{C}_2} \phi(x,y,\theta)\,dx\,dy\,d\theta, \tag{7}$$

giving in the case of k C's,

$$p(x_1,y_1,x_2,y_2) = q^k(x_1,y_1,x_2,y_2) \quad \text{and} \quad p(x_1,y_1,x_2,y_2) = G\{q(x_1,y_1,x_2,y_2)\}$$

in the general case. Thus the 2nd moment

$$\mu_2'(Y) = \int_A\int\int_A\int q^k(x_1, y_1, x_2, y_2)\,dx_1dy_1dx_2dy_2 \quad \text{for } k\ C\text{'s} \tag{8}$$

and

$$\mu_2'(Y) = \int_A\int\int_A\int G\{q(x_1, y_1, x_2, y_2)\}\,dx_1dy_1dx_2dy_2 \tag{9}$$

in the general case. In general, for the mth moment, the probability that $(x_1, y_1) \dots (x_m, y_m)$ are not covered by k C's is

$$q^k(x_1, y_1, x_2, y_2, \dots, x_m, y_m),$$

where

$$q(x_1, y_1, x_2, y_2, \dots, x_m, y_m) = \int_0^{2\pi}\int\int_{T-\bar{C}_1-\bar{C}_2-\dots-\bar{C}_m}\phi(x, y, \theta)\,dx\,dy\,d\theta, \tag{10}$$

and $T-\bar{C}_1-\bar{C}_2\dots-\bar{C}_m$ is the area of T outside C's centred at (x_1, y_1), (x_2, y_2), $\dots$, (x_m, y_m) and rotated through 180°.

Thus the mth moment is equal to

$$\left.\begin{aligned} \mu_m'(Y) &= \int_A\int\dots\int_A\int q^k(x_1, y_1, x_2, y_2, \dots, x_m, y_m)\,dx_1dy_1dx_2dy_2\dots dx_mdy_m \\ &\text{in the case of } k\ C\text{'s, or} \\ &= \int_A\int\dots\int_A\int G\{q(x_1, y_1, x_2, y_2, \dots, x_m, y_m)\}\,dx_1dy_1dx_2dy_2\dots dx_mdy_m \end{aligned}\right\} \tag{11}$$

in the general case.

4. Uniform distribution of covering areas at constant inclination

As mentioned above, in the cases with which we shall be dealing, the areas C are equally likely to fall anywhere over T, and the angle θ is constant. The function $\phi(x, y, \theta)$ can be put equal to $1/T$ for points of T and zero outside; the variable θ, and integration with respect to it, may be omitted.

The function $q(x_1, y_1)$ is the fraction of T not covered by a C centred at (x_1, y_1) and rotated through 180°, and in general $q(x_1, y_1, x_2, y_2, \dots, x_m, y_m)$ is the fraction of T outside m C's centred at (x_1, y_1), (x_2, y_2), $\dots$, (x_m, y_m) and rotated through 180°.

Instead of the variate Y we can consider Y/A, i.e. the fraction of A not covered by k C's, and to obtain its mth moment we divide $\mu_m'(Y)$ by A^m. Also the quantity $dx_1dy_1\dots dx_mdy_m/A^m$ is the probability of obtaining m centres in the elements of area $dx_1, dy_1, \dots, dx_mdy_m$ of A if these centres are uniformly distributed over A.

We thus obtain the following result from (11): *the mth moment of the fraction of A not covered by k C's with their centres falling at random on T is equal to the kth moment of the fraction of T not covered by m C's with their centres falling at random on A and rotated through* 180°.

In the case $k = m = 1$ we can express this in a slightly different way if we (i) deal with the area common to the two areas concerned, (ii) regard all orientations as possible and as equally likely, and (iii) deal with areas rather than fractions. We obtain, in fact, the following result: *the integral of the overlap of C and A, when the centre of C is taken over T and all orientations are permitted, is equal to the corresponding integral of the overlap of C and T, for all positions of the centre of C on A and for all orientations.*

In the practical cases with which we shall deal, the area A is always 'well inside' T, i.e. every point of A can be reached by a C centred somewhere in T. In such cases the formula

for the mean overlap is simple; we have $m = 1$ and the fraction of T not covered by one C is $(T-C)/T$, which is constant for all (x_1, y_1), so that its kth moment is $(T-C)^k/T^k$, i.e.

$$\mu_1'(Y/A) = \left(\frac{T-C}{T}\right)^k. \tag{12}$$

If the number of C's follows a probability generating function $G(u)$, the mean is given by

$$\mu_1'(Y/A) = G\left(\frac{T-C}{T}\right). \tag{13}$$

For the 2nd moment we are concerned with two C's centred at (x_1, y_1) and (x_2, y_2), and if their common area is $\Omega(x_1, y_1, x_2, y_2)$, we have

$$q(x_1, y_1, x_2, y_2) = \frac{T-2C+\Omega(x_1, y_1, x_2, y_2)}{T}. \tag{14}$$

The 2nd moment $\mu_2'(Y/A)$ is then the expectation of q^k or $G(q)$ for all pairs of points over A, and we no longer have a simple formula as in the case of the 1st moment. The overlap Ω, however, depends on the relative positions of the two C's, and therefore the number of variables in the integration is reduced from 4 to 2 or 1. This is illustrated in the following examples.

5. Circles falling on a fixed square

Assume A to be a square of unit side (i.e. $A = 1$), C a circle radius a and T a 'square with rounded corners', whose boundary is at a distance a outside the sides of A. Thus

$$T = 1 + 4a + \pi a^2 \tag{15}$$

and

$$C = \pi a^2. \tag{16}$$

It is seen that the fraction q of T outside the two circles centres (x_1, y_1) and (x_2, y_2) is a function only of the distance r between these points, and can therefore be written as $q(r)$. Hence if $\phi(r)$ is the frequency function of r, we obtain

$$\mu_2'(Y) = \int_0^{\sqrt{2}} q^k(r)\,\phi(r)\,dr, \tag{17}$$

or

$$\mu_2'(Y) = \int_0^{\sqrt{2}} G\{q(r)\}\,\phi(r)\,dr. \tag{18}$$

The area $\Omega(r)$ common to two circles radii a with centres distant r apart is

$$\Omega(r) = 2a^2(\theta - \sin\theta\cos\theta), \tag{19}$$

where

$$r = 2a\cos\theta \quad (r \leqslant 2a), \tag{20}$$

and

$$\Omega(r) = 0 \qquad (r \geqslant 2a), \tag{21}$$

and

$$q(r) = \frac{1+4a-\pi a^2+\Omega(r)}{1+4a+\pi a^2}, \tag{22}$$

where $\Omega(r)$ is given by (19), (20) and (21).

To obtain the frequency function $\phi(r)$ of r, we note that

$$r^2 = (x_1-x_2)^2 + (y_1-y_2)^2, \tag{23}$$

where x_1, x_2, y_1 and y_2 are uniformly and independently distributed in the range 0, 1. The difference $\xi = |x_1 - x_2|$ follows the 'triangular' distribution

$$df = 2(1-\xi)\,d\xi, \tag{24}$$

from $\xi = 0$ to 1, so that the quantity

$$u = \xi^2 = (x_1-x_2)^2$$

follows the distribution

$$df = \frac{1-\sqrt{u}}{\sqrt{u}}\,du. \tag{25}$$

Similarly, $v = (y_1 - y_2)^2$ follows independently the same distribution

$$df = \frac{1-\sqrt{v}}{\sqrt{v}}\,dv. \tag{26}$$

The distribution of $$r = \sqrt{(u+v)}$$

is obtained by integrating the product

$$\frac{\sqrt{(1-u)}\sqrt{(1-v)}}{\sqrt{(uv)}} \tag{27}$$

of the frequencies of u and v over that part of the line $u+v = r^2$ within the square of unit side in which the point u, v can lie, and we obtain without much difficulty for the frequency function of r,

$$\phi(r) = 2r(\pi - 4r + r^2) \quad \text{for} \quad 0 < r < 1, \tag{28}$$

and $$\phi(r) = 2r(4\sin^{-1} 1/r + 4\sqrt{(r^2-1)} - r^2 - \pi - 2) \quad \text{for} \quad 1 < r < \sqrt{2}. \tag{29}$$

Thus the 2nd moment of the fraction of the unit square uncovered is given by the integral (17) or (18), where $q(r)$ is given by (22), (19), (20) and (21) and $\phi(r)$ is given by (28) and (29).

It does not appear possible to reduce the integral (17) simply to elementary functions, and quadrature must be used. The integrand has discontinuities in its first derivative at $r = 1$ and $r = 2a$, so that the integration must be carried out separately over the intervals with these as end-points.

6. Overlap of circles on fixed rectangle*

We replace the square A of the previous section by a rectangle A; for convenience we assume its sides to be $\sqrt{b}$ and $1/\sqrt{b}$, where $b > 1$, so that the ratio of the longer to the shorter side is b and the area is unity. The centres of the circles radii a are assumed to be equally likely to fall anywhere in a 'rectangle with rounded corners' T, whose boundary is at a distance a outside A, i.e.

$$T = 1 + \pi a^2 + 2a(\sqrt{b} + 1/\sqrt{b}). \tag{30}$$

To obtain the 2nd moment of the fraction of the area of A not covered, we calculate an integral similar to (17) or (18). The function

$$q(r) = \frac{T - 2\pi a^2 + \Omega(r)}{T}$$

is derived from $\Omega(r)$, which remains the same, but the frequency distribution $\phi(r)$, the distance between a pair of points chosen at random in the rectangle, is different.

The co-ordinates x_1 and x_2 are uniformly and independently distributed in the range O, $\sqrt{b}$ (if we take Ox parallel to the longer side). The distribution of $u = (x_1 - x_2)^2$ is thus seen from (25) to be

$$df = \frac{1-\sqrt{(u/b)}}{\sqrt{(u/b)}}\frac{du}{b}$$

$$= \frac{\sqrt{b}-\sqrt{u}}{b\sqrt{u}}\,du, \tag{31}$$

while the distribution of $v = (y_1 - y_2)^2$ is

$$df = \frac{1-\sqrt{(bv)}}{\sqrt{(bv)}}\,b\,dv. \tag{32}$$

* This problem was solved by Robbins (1945); see footnote on p. 1.

The distribution of $r = \sqrt{(u+v)}$ is obtained by integrating the product

$$\frac{(\sqrt{b}-\sqrt{u})(1-\sqrt{(bv)})}{\sqrt{(buv)}}$$

over that part of the line $u+v = r^2$ within the rectangle $0 \leqslant u \leqslant b$, $0 \leqslant v \leqslant 1/b$. This gives

$$\phi(r) = \phi_1(r) = 2r[\pi - 2r(\sqrt{b}+\sqrt{(1/b)})+r^2] \quad \text{for} \quad r < 1/\sqrt{b}, \tag{33}$$

$$\left.\begin{aligned} &\phi(r) = \phi_2(r) = 2r[2\alpha - 1/b - 2r\sqrt{b}(1-\cos\alpha)] \quad \text{for} \quad 1/\sqrt{b} < r < \sqrt{b}, \\ &\text{where} \qquad \alpha = \sin^{-1} 1/r\sqrt{b}, \end{aligned}\right\} \tag{34}$$

$$\left.\begin{aligned} &\text{and} \quad \phi(r) = \phi_3(r) = 2r[2(\alpha-\beta) - b - 1/b + 2r\sin\beta/\sqrt{b} + 2r\sqrt{b}\cos\alpha - r^2] \\ &\text{for} \qquad \sqrt{b} < r < \sqrt{(b+1/b)}, \\ &\text{where} \qquad \beta = \cos^{-1}\sqrt{b}/r. \end{aligned}\right\} \tag{35}$$

Thus the 2nd moment of Y can be found from (19)–(21) together with (30) and (33)–(35).

7. Overlap of rectangles on a fixed rectangle

Assume that the fixed rectangle A has sides a and b and the covering rectangles C have sides α and β. The latter are assumed to be dropped with sides α parallel to the side a and with their centres anywhere inside the rectangle T, which is concentric with ab and has sides $a+\alpha$ and $b+\beta$. To calculate the 2nd moment of the fraction Y/A of A not covered by k C's, we use (18) and calculate the expectation of $q(x_1, y_1, x_2, y_2)$, the fraction of T not covered by two C's with their centres (x_1, y_1) and (x_2, y_2) falling at random in A. The area common to two C's is readily seen to depend only on the difference ξ of the x co-ordinates of their centres and on the similar difference η of their y co-ordinates. In fact, the area can be written as

$$\Omega(x_1, y_1, x_2, y_2) = [\alpha-\xi][\beta-\eta], \tag{36}$$

where the symbol* $[x]$ stands for x when $x > 0$ and is zero when $x < 0$, and we obtain

$$q(x_1, y_1, x_2, y_2) = 1 + \frac{[\alpha-\xi][\beta-\eta]-2\alpha\beta}{(a+\alpha)(b+\beta)}. \tag{37}$$

To obtain the expectation of q^k, we need the frequency distribution of ξ and η. As in § 5, ξ is readily seen to follow the frequency distribution

$$df = \frac{2(a-\xi)}{a^2}d\xi \tag{38}$$

between O and a, with a similar distributon for η, and we obtain the result

$$\mu_2'(Y/A) = \frac{4}{a^2b^2}\int_0^a\int_0^b\left\{1+\frac{[\alpha-\xi][\beta-\eta]-2\alpha\beta}{(a+\alpha)(b+\beta)}\right\}^k(a-\xi)(b-\eta)\,d\xi\,d\eta. \tag{39}$$

If the kth power be expanded, the resulting integrals are, with the exception of the first, the product of integrals whose upper limits are $a' = \min(a,\alpha)$ and $b' = \min(b,\beta)$ respectively. We obtain

$$\begin{aligned} \mu_2'(Y/A) = {} & \frac{4}{a^2b^2}\int_0^{a'}\int_0^{b'}\left\{1+\frac{(\alpha-\xi)(\beta-\eta)-2\alpha\beta}{(a+\alpha)(b+\beta)}\right\}^k(a-\xi)(b-\eta)\,d\xi\,d\eta \\ & + \frac{4}{a^2b^2}\left\{1-\frac{2\alpha\beta}{(a+\alpha)(b+\beta)}\right\}^k\left\{\int_0^a\int_0^b(a-\xi)(b-\eta)\,d\xi\,d\eta - \int_0^{a'}\int_0^{b'}(a-\xi)(b-\eta)\,d\xi\,d\eta\right\}. \end{aligned} \tag{40}$$

* The writer is indebted to Neyman & Bronowski for this convenient notation (see below).

By a simple change of variable we obtain

$$\mu_2'(Y/A) = \frac{4}{a^2b^2}\int_{[\alpha-a]}^{\alpha}\int_{[\beta-b]}^{\beta}\left\{1+\frac{uv-2\alpha\beta}{(a+\alpha)(b+\beta)}\right\}^k (u+a-\alpha)(v+b-\beta)\,du\,dv$$
$$+\frac{1}{a^2b^2}\left\{1-\frac{2\alpha\beta}{(a+\alpha)(b+\beta)}\right\}^k \{b^2[a-\alpha]^2+a^2[b-\beta]^2-[a-\alpha]^2[b-\beta]^2\}, \tag{41}$$

which is the result obtained by Bronowski & Neyman by a rather different method.*

8. Overlap of circles on a fixed circle

We now consider a fixed circle A of unit area and therefore of radius $b = 1/\sqrt{\pi}$, with k circles C of radius a dropped at random with their centres uniformly distributed over a circle T of radius $a+b$. The 2nd moment of the fraction Y/A of A not covered is, as in § 5, the expectation of $q^k(r)$, where $q(r)$ is the fraction of T not covered by two circles with centres falling at random in A a distance r apart. We have

$$q(r) = \frac{1+2a\sqrt{\pi}-\pi a^2+\Omega(r)}{1+2a\sqrt{\pi}+\pi a^2}, \tag{42}$$

where $\Omega(r)$, the overlap of two circles radius a with centres apart, is given by (19), (20) and (21) as before. We thus need the frequency distribution $\phi(r)$ of the distance between two points chosen at random in the circle of unit area to obtain

$$\mu_2'(Y/A) = \int_0^{2\sqrt{\pi}} q^k(r)\,\phi(r)\,dr. \tag{43}$$

To do this we use a fairly straightforward geometrical method, finding first the probability integral

$$F(r) = \int_0^r \phi(r)\,dr, \tag{44}$$

which is the probability that the distance between the two random points is less than r.

The probability that the first point is between v and $v+dv$ from the centre is $2v\,dv/b^2$, while if

$$A(v) = \text{area common to circles radii } b \text{ and } r \text{ with centres distance } v \text{ apart}, \tag{45}$$

it follows that the probability of the second point being within r of the first is $A(v)/\pi b^2$. Hence

$$F(r) = \int_0^b \frac{2v}{b^2}\frac{A(v)}{\pi b^2}\,dv. \tag{46}$$

Construct the triangle with sides r, b and v, and let the angles opposite to these be θ, ϕ and ψ. Then the following can be readily verified:

$$\text{If } r<b,\quad \left.\begin{aligned} A(v) &= b^2\theta+r^2\phi-br\sin\psi &&\text{if } b-r<v<b,\\ &= \pi r^2 &&\text{if } 0<v<b-r.\end{aligned}\right\} \tag{47}$$

$$\text{If } b<r<2b,\quad \left.\begin{aligned} A(v) &= b^2\theta+r^2\phi-br\sin\psi &&\text{if } r-b<v<b,\\ &= \pi b^2 &&\text{if } 0<v<r-b.\end{aligned}\right\} \tag{48}$$

The integration in (46) is carried out by parts, with ψ as the ultimate variable of integration, and to do this we obtain the result

$$A'(v) = -\frac{2br\sin\psi}{v}. \tag{49}$$

Putting

$$r = 2b\sin\tfrac{1}{2}\alpha, \tag{50}$$

we obtain, over the whole range of r,

$$F(r) = \frac{\alpha}{\pi}+r^2(\pi-\alpha)-\frac{r\cos\frac{1}{2}\alpha}{\sqrt{\pi}}-\frac{r^2\sin\alpha}{2}, \tag{51}$$

* And by Robbins (1945).

while differentiation yields the frequency distribution as

$$\phi(r) = 2r(\pi-\alpha) - \frac{\sqrt{\pi}}{2\cos\frac{1}{2}\alpha}(4r^2 - \pi r^4). \tag{52}$$

It will be noted as a matter of interest that the chance of the two random points falling further apart than the radius of the circle is $1 - F(b) = \frac{3\sqrt{3}}{4\pi}$ or 9/22 nearly.*

We thus obtain the 2nd moment of the uncovered area from (42), (43) and (52).

9. Use of probability generating functions

(i) *Binomial*

It is interesting to apply first the binomial distribution of k, the number of C's dropped uniformly and at random on T. Assume that T contains the centres of all the C's which touch or cover A, and that S is some larger area including T. If l C's are dropped at random on S, the probability generating function of the number of centres falling on T is

$$G_1(u) \equiv \left(\frac{S-T+Tu}{S}\right)^l. \tag{53}$$

Thus, from the general result of § 4, the mth moment of the fraction of A uncovered is the expectation of $\left(\frac{S-T+Tq}{S}\right)^l$, where q is the fraction of T uncovered by m C's falling on A. But the expression within brackets is the fraction of S uncovered. The use of the binomial generating function is thus verified.

(ii) *Poisson*

The Poisson distribution next suggests itself. If the number of centres follows this distribution with a mean of λ per unit area, the probability generating function is

$$G_2(u) \equiv e^{\lambda T(u-1)} \tag{54}$$

and the mth moment will be expectation over A of

$$e^{\lambda T(q-1)}.$$

Alternatively, we could write this as

$$\mu'_m(Y/A) = \exp(e^{-\lambda Z}), \tag{55}$$

where Z = area of overlap of m C's falling on A. In particular,

$$\text{mean value of } Y/A = \mu'_1(Y/A) = e^{-\lambda C}. \tag{56}$$

Thus the mth moment of Y/A is related to the characteristic function of Z, but this result does not appear to be of any theoretical importance: it does not, for instance, throw any light on the frequency distribution of Y/A. Formula (55) does, however, demonstrate the fact, which is otherwise obvious, that the area T does not enter into the frequency distribution of the fraction of A not covered by C's whose fall follows the Poisson distribution.

As far as the calculation of the variance is concerned, we need to calculate first the 2nd moment of $e^{-\lambda Z}$. In the cases where the falling areas are circles, the area of overlap Z of two circles is equal to $2C - \Omega(r)$, where $\Omega(r)$ is the function given above ((19) etc.) for the area common to two C's with centres r apart. The 2nd moment is thus

$$e^{-2\lambda C}\int e^{\lambda\Omega(r)}\phi(r)\,dr, \tag{57}$$

where $\phi(r)$ is the frequency function of r, the formulae for which are given above for the various cases.

* The solution to this problem (no. 698), given by Whitworth (1897), contains an error, resulting in the incorrect value of 35/88 nearly.

In the case of rectangles falling on rectangles, the necessary formula for the variance is given by Neyman & Bronowski in the form of a series, viz.

$$\mu_2(Y/A) = \frac{4e^{-2\alpha\beta\lambda}}{a^2b^2}\sum_{s=1}^{\infty}\frac{(\lambda\alpha\beta)^s\alpha\beta}{s!\,(s+1)^2\,(s+2)^2}$$
$$\times\{(s+2)\,a-\alpha+[\alpha-a]\,(1-a/\alpha)^{s+1}\}\{(s+2)\,b-\beta+[\beta-b]\,(1-b/\beta)^{s+1}\}. \quad (58)$$

Table 1. *Variance of fraction of fixed area not covered by areas C falling according to Poisson distribution*

(i) Circles falling on square; (ii) circles falling on rectangle 2×1; (iii) circles falling on rectangle 4×1; (iv) circles falling on circle; (v) squares falling on square (sides parallel).

Size of falling area $C \div$ size of fixed area	Case	Mean area not covered 0·25	0·50	0·75
		Variance of area not covered		
0·2	(i)	0·0186	0·0303	0·0254
	(ii)	0·0182	0·0296	0·0248
	(iii)	0·0169	0·0275	0·0229
	(iv)	0·0190	0·0310	0·0260
	(v)	0·0182	0·0300	0·0252
1·0	(i)	0·0608	0·0964	0·0789
	(ii)	0·0573	0·0904	0·0743
	(iii)	0·0489	0·0766	0·0627
	(iv)	0·0620	0·0983	0·0810
	(v)	0·0593	0·0945	0·0781
1·8	(i)	0·0816	0·1262	0·1026
	(ii)	0·0771	0·1193	0·1040
	(iii)	0·0657	0·1015	0·0824
	(iv)	0·0829	0·1281	0·1040
	(v)	0·0790	0·1230	0·1002

In general, the variance increases in the following order as between the different combinations of shapes:

(1) Circles on rectangle 4×1, (iii).
(2) Circles on rectangle 2×1, (ii).
(3) Squares on square, (v).
(4) Circles on square, (i).
(5) Circles on circle, (iv).

There is an exception in the last case considered, however (mean fraction of area not covered $=0{\cdot}75$, size of falling area $\div$ fixed area $=1{\cdot}8$), when the order is changed somewhat. However, the variances are generally of the same approximate magnitude.

(iii) *Contagious distribution*

Neyman & Bronowski have included in their study the case of a contagious law of type A with two parameters (see Neyman, 1939). Here the probability generating function is

$$G_3(u) = e^{m\{e^{\lambda T(u-1)}-1\}}. \quad (59)$$

They have pointed out that the expression of this as a series enables calculations to be utilized from the Poisson distribution.

In the general case the sth moment of the fraction of A not covered is the expectation of

$$e^{m(e^{-\lambda Z}-1)}, \quad (60)$$

where Z is the area common to s C's falling with their centres on A.

10. NUMERICAL RESULTS

It is impossible to calculate complete tables covering all cases, but it is of interest to calculate a few values for the purpose of comparison, and the following combinations of areas have been considered:

	Fixed area	Falling areas
(i)	Square	Circles
(ii)	Rectangle 2×1	Circles
(iii)	Rectangle 4×1	Circles
(iv)	Circle	Circles
(v)	Square	Squares (with sides parallel to fixed square)

In each case the fixed area has been made of unit size, while the falling areas were respectively $C = 0{\cdot}2$, $1{\cdot}0$ and $1{\cdot}8$. The areas were assumed to fall according to the Poisson distribution, the number of centres per unit area being such that the expected fraction of the fixed area A not covered was respectively $0{\cdot}25$, $0{\cdot}5$ and $0{\cdot}75$. Since from equation (56) the expected fraction of A not covered is $m = e^{-\lambda C}$, the relations between λ and C for the 9 combinations of m and C are

$$\lambda C = \log_e 4, \quad \log_e 2 \quad \text{and} \quad \log_e 4/3.$$

The 2nd moments were determined by quadrature from formula (57), where the falling areas were circles, and by direct evaluation of the series (58) for the case of squares on squares. The results are given in Table 1.

11. EXPERIMENTAL INVESTIGATION

Before the work of Robbins and Neyman & Bronowski was brought to the notice of the writer, an attempt was made to obtain experimentally a general formula for the variance of the fraction of area not covered. Attention was confined to the case of circles falling on squares, the centres of the former being chosen randomly (by means of random numbers) within the area T whose boundary is at a distance a outside the sides of the unit square A.

For each combination of C and k, samples of up to 200 in size were drawn. The various combinations were as given in Table 2.

Table 2. *Ranges of k and C covered in experimental determination of variance*

$\dfrac{\text{Area of circles dropped}}{\text{Area of square}} = C$	No. of circles $= k$
0·0077	5, 10, 15
0·031	5, 10, 15, 20, 30
0·033	5, 20, 40, 80, 120
0·25	1, 2, 4, 6
1	1, 2, 4, 6

Three methods were used to measure the fraction of the fixed square not covered in each sample. Method P involved the measurement of the covered area by planimeter. Method L utilized a photoelectric cell to measure the amount of light passing through a glass plate on which black paper disks had been stuck. Method C consisted of a simple count of squares on graduated paper, and generally this was the most convenient to operate. (The two neighbouring values of C were used to compare methods L and P.)

For each combination of C and k the average fraction not covered, $\overline{Y}$, was compared with the theoretical value

$$m = (1 - C/T)^k. \tag{61}$$

The observed standard deviation of the observations being s, the appropriate criterion for testing the mean is

$$t = \frac{\bar{Y}-m}{s/\sqrt{P}},$$

P being the number of observations in the sample. The results are given in Table 3.

Table 3. *Comparison between observed and expected values of fraction of area not covered*

P = planimeter method. L = photoelectric method. C = counting method.

Area of circle / Area of square $= C$	No. of circles k	No. in sample P	Mean area not covered		Standard deviation s	Deviation $t = \frac{\bar{Y}-m}{s/\sqrt{P}}$	Method of measurement
			Observed $\bar{Y}$	Expected m			
0·0077	5	100	0·9694	0·9683	0·0055	2·0000*	P
0·0077	10	100	0·9394	0·9376	0·0088	2·0454*	P
0·0077	15	100	0·9105	0·9079	0·0090	2·8889†	P
0·031	5	100	0·8924	0·8961	0·0222	−1·6667	P
0·031	10	200	0·8007	0·8030	0·0334	−0·9739	P
0·031	15	100	0·7134	0·7196	0·0404	−1·5347	P
0·031	20	100	0·6349	0·6449	0·0423	−2·3641*	P
0·031	30	100	0·5057	0·5179	0·0493	−2·4746*	P
0·033	5	100	0·8853	0·8901	0·0250	−5·9200†	L
0·033	20	100	0·6326	0·6278	0·0500	0·9600	L
0·033	40	100	0·3969	0·3941	0·0367	0·7629	L
0·033	80	75	0·1617	0·1553	0·0358	1·4756	L and P
0·033	120	50	0·0657	0·0612	0·0255	1·2478	P
0·25	1	30	0·8916	0·8950	0·0824	−0·2260	C
0·25	2	110	0·7853	0·8009	0·1214	−1·3477	C
0·25	4	110	0·6263	0·6415	0·1481	−1·0764	C
0·25	6	110	0·4891	0·5138	0·1307	−1·9820*	C
1·0	1	20	0·7583	0·7653	0·2350	−0·1332	C
1·0	2	50	0·5890	0·5856	0·2346	0·1025	C
1·0	4	50	0·3861	0·3430	0·2334	1·3057	C
1·0	6	50	0·1686	0·2008	0·1630	−1·3968	C

* Between 5 and 1 % levels. † Beyond 1 % level.

An examination of the values of t shows that too many of them are outside the 5 % significance levels, while in each set corresponding to one value of C the values are too frequently of the same sign. The worst deviation is for $C = 0{\cdot}033$ and $k = 5$, with $t = -5{\cdot}92$, but this only corresponds to a difference between the observed mean of 0·885 and the theoretical mean of 0·890. The test is thus very sensitive and the deviations are not serious, and they arise from imperfections in the technique which have not been investigated in detail.

As regards random errors of measurement as distinct from bias, it was not possible to carry out a systematic estimation of the contribution of this source to the total variation. A series of repeated measurements for the case $C = 0{\cdot}031$, $k = 30$, for which the observed mean was 0·51, showed that the individual measurements had a standard error of about 0·009. As the total standard deviation in this case was 0·049, the true estimate of the standard error (i.e. omitting the error of measurement) was $\sqrt{(0{\cdot}049^2 - 0{\cdot}009^2)} = 0{\cdot}048$, and for our purposes this difference is negligible. There is thus some evidence for assuming that this method of estimating the variance was satisfactory.

12. Derivation of empirical formula for the variance

The consideration that the theoretical variance σ^2 of the fraction uncovered must be small whenever the mean m of this fraction is near the limits of its range, zero or unity, suggests that we might try the relation

$$\sigma^2 \sim m(1-m)$$

Table 4. *Derivation of empirical formula*

$\frac{\text{Area of circles}}{\text{Area of square}} = C$	No. of circles k	Ratio T/C	Observed variance $= s^2$	$g = \frac{s^2}{m(1-m)}$	Mean values of g $= g(C)$	$(T/C)^{3/2} \times g(C)$	Empirical value of variance $= \frac{2{\cdot}3m(1-m)}{(T/C)^{3/2}} = \sigma_1^2$	Exact value of variance $= \sigma^2$	Percentage error in σ_1^2	Percentage error in s^2
0·0077	5	155·7	0·0000303	0·000973			0·0000363	—	—	—
0·0077	10	155·7	0·0000774	0·001296	0·00108	2·10	0·0000692	—	—	—
0·0077	15	155·7	0·0000810	0·000980			0·0000990	—	—	—
0·031	5	46·1	0·000493	0·00530			0·000684	—	—	—
0·031	10	46·1	0·00112	0·00704			0·00116	—	—	—
0·031	15	46·1	0·00163	0·00810	0·00759	2·38	0·00148	—	—	—
0·031	20	46·1	0·00179	0·00780			0·00168	—	—	—
0·031	30	46·1	0·00243	0·00972			0·00184	—	—	—
0·033	5	43·5	0·000625	0·00638			0·000785	—	—	—
0·033	20	43·5	0·00250	0·01071			0·00188	—	—	—
0·033	40	43·5	0·00135	0·00564	0·00876	2·51	0·00192	—	—	—
0·033	80	43·5	0·00128	0·00976			0·00105	—	—	—
0·033	120	43·5	0·000650	0·01132			0·000461	—	—	—
0·25	1	9·5	0·00679	0·0723			0·00736	0·00731	0·7	− 7·1
0·25	2	9·5	0·0147	0·0924	0·0820	2·41	0·01249	0·0123	2·0	19·5
0·25	4	9·5	0·0219	0·0954			0·0180	0·0176	2·2	24·4
0·25	6	9·5	0·0171	0·0681			0·0196	0·0185	5·9	− 7·6
1·0	1	4·3	0·0552	0·307			0·0471	0·0508	−7·3	8·7
1·0	2	4·3	0·0550	0·223	0·234	2·05	0·0636	0·0651	−2·3	−15·5
1·0	4	4·3	0·0545	0·241			0·0590	0·0540	9·3	0·9
1·0	6	4·3	0·0266	0·166			0·0420	0·0339	23·9	−21·5

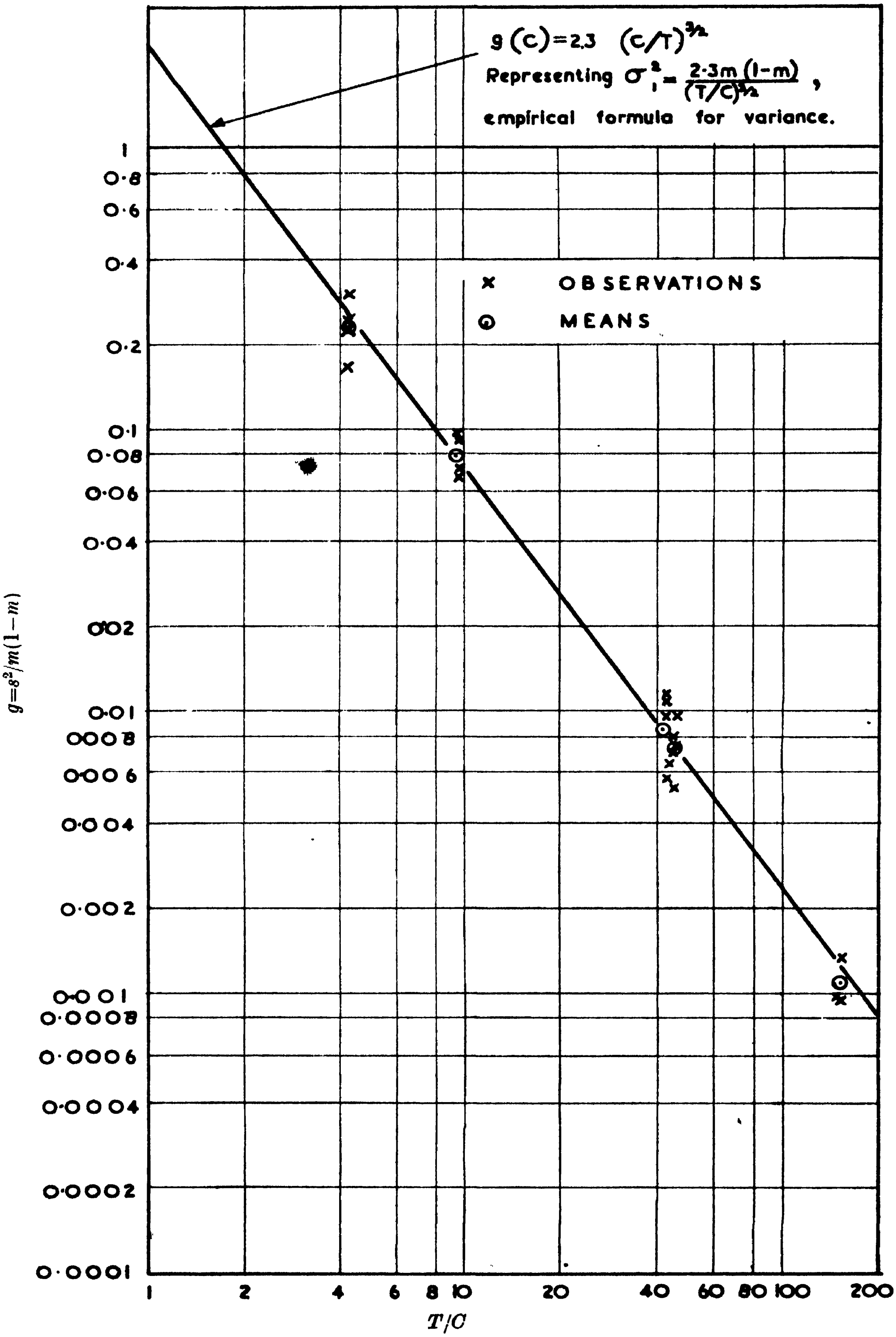


Fig. 2. Derivation of empirical formula for variance.

for given C. Accordingly we have calculated the quantity

$$g = \frac{s^2}{m(1-m)}, \tag{62}$$

and the results are given in Table 4.

For each value of C the values of g are by no means constant, but the variation is not excessive, and it is considered that for practical purposes we can take g to be a function of C, at least over the range considered. To obtain a suitable form for this function, it was decided, for very general reasons, to seek a simple relation between $g(C)$ and T/C, the latter being roughly the number of C's which could be placed on T if they could be fitted together without overlapping.

Table 5. *Comparison of empirical formula for variance with exact value*

σ^2 = exact value, σ_1^2 = empirical value, E = percentage error = $100\,(\sigma_1^2-\sigma^2)/\sigma^2$.

πa^2		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
0·2	σ^2	0·00501	0·00873	0·0114	0·0132	0·0144	0·0150
	σ_1^2	0·00516	0·00896	0·0117	0·0135	0·0147	0·0154
	E	+3·0	+2·6	+2·6	+2·3	+2·1	+2·7
0·4	σ^2	0·0156	0·0246	0·0290	0·0305	0·0300	0·0274
	σ_1^2	0·0149	0·0237	0·0284	0·0304	0·0305	0·0294
	E	−4·5	−3·7	−2·1	−0·3	+1·7	+7·3
0·6	σ^2	0·0277	0·0403	0·0447	0·0427	0·0395	0·0339
	σ_1^2	0·0257	0·0384	0·0431	0·0433	0·0407	0·0370
	E	−7·2	−4·7	−3·6	+1·4	+3·0	+9·1
0·8	σ^2	0·0398	0·0541	0·0552	0·0501	0·0428	0·0351
	σ_1^2	0·0365	0·0517	0·0551	0·0525	0·0471	0·0407
	E	−8·3	−4·4	−0·2	+4·8	+10·0	+16·0
1·0	σ^2	0·0508	0·0651	0·0628	0·0540	0·0436	0·0339
	σ_1^2	0·0471	0·0636	0·0648	0·0590	0·0507	0·0420
	E	−7·3	−2·3	+3·2	+9·3	+16·3	+23·9
1·2	σ^2	0·0605	0·0737	0·0677	0·0554	0·0427	0·0317
	σ_1^2	0·0571	0·0740	0·0724	0·0635	0·0525	0·0419
	E	−5·6	+0·4	+6·9	+14·6	+22·9	+32·5
1·4	σ^2	0·0689	0·0804	0·0706	0·0554	0·0410	0·0292
	σ_1^2	0·0667	0·0832	0·0786	0·0665	0·0531	0·0411
	E	−3·2	+3·5	+11·3	+20·0	+29·5	+40·7
1·6	σ^2	0·0763	0·0855	0·0723	0·0546	0·0389	0·0267
	σ_1^2	0·0757	0·0913	0·0834	0·0684	0·0530	0·0398
	E	−0·8	+6·8	+15·4	+25·3	+36·3	+49·1
1·8	σ^2	0·0828	0·0895	0·0730	0·0531	0·0367	0·0244
	σ_1^2	0·0843	0·0985	0·0873	0·0695	0·0524	0·0383
	E	+1·8	+10·1	+19·6	+30·9	+42·8	+57·0

Fig. 2 shows the result of plotting the observed mean value of $g(C)$ against T/C on logarithmic scales. The points (the values of $s^2/m(1-m)$ for various values of k are plotted in addition to the mean), lie reasonably close to a straight line of slope $-1{\cdot}5$, indicating the relation

$$g(C) \sim (C/T)^{\frac{3}{2}}. \tag{63}$$

The values of $(T/C)^{\frac{3}{2}}g(C)$ are shown in Table 4, where the values are seen to lie between 2 and 2·5 with an average of 2·3. Thus we derive the rough empirical formula

$$\left.\begin{aligned} \sigma_1^2 &= \frac{2{\cdot}3m(1-m)}{(T/C)^{\frac{3}{2}}}, \\ \text{where}\quad m &= (1-C/T). \end{aligned}\right\} \tag{64}$$

These are given in Table 4, together with the values in some cases of the percentage error of σ_1^2 compared with the true value σ^2 obtained from the methods described in § 5; the percentage error in the estimate s^2 observed experimentally is also given. (It was not possible to evaluate σ^2 in all cases, as the computation is somewhat laborious.) Another set of comparisons between the empirical and the exact formula is given below in Table 5, over the range $C = \pi a^2$ from 0·2 to 1·8 and $k = 1, 2, 3, 4, 5, 6$.

It will be seen that the empirical formula gives quite a satisfactory fit, e.g. with an error less than 10 %, over a considerable part of the range studied, but that the error tends to increase, i.e. the formula exaggerates the variance, as C and k increase.

13. Use of the empirical formula for the Poisson case

If k circles are dropped with their centres falling at random on the area T the mean area not covered can be written as

$$m(k) = (1 - C/T)^k,$$

and the empirical formula for the variance as

$$\mu_2(k) = \frac{2\cdot 3m(k)\,[1-m(k)]}{(T/C)^{\frac{3}{4}}}.$$

Table 6. *Comparison of empirical formula for variance of area not covered in case of circles falling on square according to Poisson distribution*

Size of falling area C ÷ fixed area		Mean area not covered, m		
		0·25	0·5	0·75
0·2	σ^2	0·0186	0·0303	0·0254
	σ_1^2	0·0196	0·0308	0·0257
	% error	5·4	1·7	1·2
1·0	σ^2	0·0608	0·0964	0·0789
	σ_1^2	0·0669	0·0981	0·0781
	% error	10·0	1·8	−1·0
1·8	σ^2	0·0816	0·1262	0·1026
	σ_1^2	0·0942	0·1348	0·1057
	% error	15·4	6·8	3·0

Hence if k follows the Poisson distribution with expectation λT, the total variance based on the empirical formula is

$$\sigma_1^2 = \sum_{k=0}^{\infty} \frac{e^{-\lambda T}(\lambda T)^k}{k!}[m^2(k) + \mu_2(k)] - m^2,$$

where

$$m = \text{expected area not covered} = e^{-\lambda C}. \tag{65}$$

We find after expanding $m^2(k)$ that

$$\sigma_1^2 = m^2\left[\left\{1 - \frac{2\cdot 3}{(T/C)^{\frac{3}{4}}}\right\} m^{-C/T} - 1\right] + \frac{2\cdot 3m}{(T/C)^{\frac{3}{4}}}. \tag{66}$$

This formula is compared with the exact values in Table 6 over the same range as in Table 1.

The agreement is again reasonably satisfactory over the greater part of the range, large positive errors occurring for large values of C and small values of m. These errors might be reduced by using a constant rather smaller than 2·3 in the empirical formula, but the point has not been investigated further.

SUMMARY

The mathematical study of bombing has given rise to the following problem. A fixed outline, such as a square or circle, is drawn on a plane, and other similar outlines are dropped at random on it. Estimates are then required of the variance of the fixed area which is not covered. Work by Robbins enables a theoretical formula to be derived, and Bronowski & Neyman have treated, by an independent method, the special case of rectangles falling on rectangles.

It is shown that in the case of circles falling on circles, squares or rectangles, the variance can be expressed as the integral with respect to r of the product of two functions, one being a simple function of the area of overlap of two circles with centres r apart, and the other being the frequency function of the distance r between two points chosen at random in the 'covered' area. This applies both to the case where the number of falling areas is fixed and where it follows a Poisson distribution. Numerical values have been calculated for a number of cases. An experimental method had been carried out prior to the above theoretical work, and the following empirical formula was derived for the variance of the fraction of a fixed square not covered by k circles, area C, falling at random on an area T containing centres of all C's which cover or touch the fixed square:

$$\sigma^2 = \frac{2{\cdot}3m(1-m)}{(T/C)^{\frac{3}{2}}},$$

where m = mean fraction of area not covered $= (1 - C/T)^k$.

This formula, and its extension to the Poisson case, have been shown to be in reasonable agreement with the exact values over a considerable range.

The writer is indebted to Miss G. O. Jeffcoate for valuable assistance in the computing and experimental work.

REFERENCES

BRONOWSKI, J. & NEYMAN, J. (1945). *Ann. Math. Statist.* **16**, 330.

NEYMAN, J. (1939). *Ann. Math. Statist.* **10**, 35.

ROBBINS, H. E. (1944). *Ann. Math. Statist.* **15**, 70.

ROBBINS, H. E. (1945). *Ann. Math. Statist.* **16**, 342.

WHITWORTH, W. A. (1897). *DCC Exercises in Choice and Chance.* Cambridge.

A STUDY OF A FIRST DYNASTY SERIES OF EGYPTIAN SKULLS FROM SAKKARA AND OF AN ELEVENTH DYNASTY SERIES FROM THEBES

By A. BATRAWI, Ph.D. and G. M. MORANT, D.Sc.

1. *Introduction*. This paper deals with forty-four male crania of 1st dynasty date (*c*. 3400 B.C.) discovered at Sakkara by Macramallah Effendi, who has published a report on the excavations (1940), and with fifty-five crania of 11th dynasty (*c*. 2000 B.C.) soldiers unearthed at Thebes in 1927 by the Egyptian Expedition of the Metropolitan Museum of Art, New York (Winlock, 1928). The cemetery at Sakkara, 20 miles south of Gizeh, was used by the middle classes of the local community. Prof. D. E. Derry has kindly provided the following notes on it:

The 1st dynasty cemetery at Sakkara excavated by Macramallah Effendi is of special interest. Comparatively few cemeteries of this date have been found, and, while the total number of forty-four skulls from which reliable measurements could be taken was small, yet the results yielded by these are such as to show that we are dealing with a race which differs in important features from those exhibited by the so-called predynastic people.

The observation that there were two races in Egypt in the early dynastic period was first made in the year 1909, when the results of measurements obtained from a series of male and female skulls of the 4th and 5th dynasties from the great necropolis surrounding the pyramids of Giza came to be examined and compared with crania from early predynastic graves. Until then the theory of an unbroken evolution of the Egyptian race from prehistoric times right through the dynastic period had been taught. It now became obvious that the culture which we know of as peculiarly Egyptian was associated with a race which could not have been derived from the predynastic people. The introduction of stone-working resulting in the erection of great tombs and statuary, as well as beautifully executed reliefs, paintings and above all writing, all pointed to a race far in advance of the predynastic people, who although skilled in the making of bowls and vases in stone as well as in pottery, and who had already attained to the discovery of the uses of copper, were, nevertheless, little removed from the Neolithic period.

The cemetery is unusual in consisting entirely of males. In the note on the skulls published in Macramallah Effendi's report it is stated that there were some females included in the collection. After the report had gone to press Macramallah Effendi informed the writer that a part of the cemetery was of 18th dynasty date. It turned out that all the female skulls came from this part and that therefore the 1st dynasty cemetery contained only remains of males. Dr Batrawi's examination of the figures confirms the statement made at the beginning of this note and shows the closeness of the relationship of the people of the 1st dynasty at Sakkara with those of the 4th and 5th dynasties from Deshasheh and Medum.

In his report on the discovery of the series of 11th dynasty skeletons at Thebes Mr H. E. Winlock (1928) says that they were found in 'a tomb in the row where the grandees of Mentuhotep's court had been buried'. He remarks:

Obviously what we had found was a soldiers' tomb. To judge from the cheapness of their burial they were only soldiers of the rank and file, and yet they had been given a catacomb presumably prepared for the dependents of the royal household, next to the tomb of the chancellor Khety. Clearly that was an especial honour. If we are right in supposing that all had been buried at once, they must have been slain in a single battle.

Prof. Derry examined the bodies on the spot, and he took measurements of the crania and of some of the long bones. About sixty bodies were counted and all proved to be those

of adult males who had died in the prime of life. Prof. Derry says that the skeletons were reburied after they had been examined.

2. *The measurements of the crania.* The Sakkara series was sexed and measured by Prof. Derry and we are indebted to him for allowing us to use his records in this paper. All the absolute measurements, given in Appendix II, are his readings with the exception of those of the *foramen magnum*, which were taken by one of the writers (A.B.) of this report. The measurements of the Thebes series were also kindly provided by Prof. Derry, together with means he had calculated. The readings for individual crania are given in Appendix III.

The technique of measurement followed by Prof. Derry is that of the Monaco Congress (Duckworth, 1913). He had used this when measuring the predynastic Egyptian series of skulls from Badari, of which part was remeasured later in London by Miss B. N. Stoessiger (1927), who followed the biometric technique. The two sets of measurements of the same fifty-three specimens have been compared (Morant, 1935), thus showing in detail what relations are to be expected between readings obtained by following the two techniques. These results were taken into account in preparing the definitions of Prof. Derry's measurements given in Appendix I below. The characters are denoted as far as possible there and in the tables by the customary index letters of the biometric technique.

3. *The nature of the two series.* Mean measurements and standard deviations for the two series are given in Table 1. The longest series of Egyptian skulls measured, known as the *E* series, came from a cemetery at Giza used from the 26th–30th dynasties (Davin & Pearson, 1924). Judging from comparisons of constants for a number of cranial characters, most of the other ancient Egyptian series described exhibit almost precisely the same order of variation as the one from Giza. In general they have been found to be rather less variable than European cranial series, while there is no evidence that there was any appreciable change in the variation exhibited by Egyptian populations during the long period from early predynastic to Roman times.

The two new series are shorter than several from Egypt previously described. Counting the number of characters for which the standard deviation for one series is greater or less than the corresponding constant for the other series, the situation is:

Sakkara and Thebes: Sakkara S.D. greater for nine and less for ten characters;

Sakkara and Giza: Sakkara S.D. greater for four and less for eleven characters;

Thebes and Giza: Thebes S.D. greater for eight and less for nine characters.

This crude comparison suggests that there can have been no marked differences between the variabilities of the three populations represented. As sets of differences are considered, the limit of significance accepted may be taken considerably higher than in the case of a single difference. Suppose that there is a real distinction if two of the standard deviations differ by an amount which is 3·5 or more times its probable error. Then one significant difference is found for the Sakkara and Thebes series (*NH*, *L*, Sakkara S.D. greater, $\Delta/\text{P.E.}\Delta = 3{\cdot}8$), none for the Sakkara and Giza, and three for the Thebes and Giza series (H' 4·1, S_1 3·5, S_2 3·5, Giza S.D. the greater in all three cases). The two new series are too short to give reliable comparisons, but the evidence suggests that the populations they represent were equally homogeneous, while both were rather less mixed in racial composition than the 26th–30th dynasty population of Giza.

4. *Comparisons of mean measurements.* Following biometric practice, it may be supposed in such a case that no statistical analysis of the series can reveal its racial components. The relationships of the series have to be judged by comparing them as wholes, on the basis of mean measurements, with other series known to exhibit unexceptional variation. It was shown by Morant (1925) that the recorded series of ancient Egyptian skulls can be divided into two groups. These were called, for convenience, the Upper and Lower Egyptian,

Table 1. *Means and standard deviations (with probable errors) of the Sakkara 1st dynasty and Thebes 11th dynasty series of male skulls*

Character*	Means		Standard deviations	
	Sakkara	Thebes	Sakkara	Thebes
L	186·9 ± 0·56 (41)	181·8 ± 0·53 (54)	5·31 ± 0·40	5·75 ± 0·37
B	138·7 ± 0·41 (43)	138·3 ± 0·41 (54)	3·99 ± 0·29	4·52 ± 0·29
B'	96·5 ± 0·39 (36)	93·6 ± 0·43 (55)	3·48 ± 0·28	4·72 ± 0·30
H'	135·4 ± 0·67 (32)	137·1 ± 0·37 (51)	5·63 ± 0·47	3·91 ± 0·26
[Aur. ht.]	114·8 ± 0·67 (27)	—	5·20 ± 0·48	—
LB	102·7 ± 0·57 (29)	100·7 ± 0·34 (46)	4·57 ± 0·40	3·40 ± 0·24
U	518·8 ± 1·5 (29)	507·4 ± 1·3 (50)	12·0 ± 1·1	13·2 ± 0·89
S_1	—	125·7 ± 0·47 (52)	—	5·01 ± 0·33
S_2	—	129·4 ± 0·56 (53)	—	6·00 ± 0·39
S_3	—	115·2 ± 0·82 (51)	—	8·63 ± 0·58
S	—	370·5 ± 1·2 (50)	—	12·7 ± 0·86
[Broca's Q']	—	300·3 ± 0·84 (49)	—	8·72 ± 0·59
fml	36·7 ± 0·26 (31)	—	2·18 ± 0·19	—
fmb	30·4 ± 0·22 (29)	—	1·74 ± 0·15	—
[$G'H$]	71·9 ± 0·55 (30)	72·0 ± 0·37 (45)	4·43 ± 0·39	3·71 ± 0·26
GB	96·5 ± 0·62 (25)	95·5 ± 0·48 (38)	4·57 ± 0·44	4·40 ± 0·34
J	127·8 (14)	127·6 ± 0·52 (32)	—	4·33 ± 0·36
[NH, L]	51·2 ± 0·50 (29)	51·8 ± 0·25 (45)	4·00 ± 0·35	2·52 ± 0·18
NB	25·4 ± 0·21 (30)	25·0 ± 0·20 (42)	1·70 ± 0·15	1·92 ± 0·14
[O_1']	38·9 ± 0·24 (26)	39·1 ± 0·16 (44)	1·84 ± 0·17	1·55 ± 0·11
[O_2]	32·5 ± 0·20 (26)	33·1 ± 0·23 (44)	1·50 ± 0·14	2·23 ± 0·16
[Prosthion GL]	99·6 ± 0·56 (26)	96·5 ± 0·47 (43)	4·23 ± 0·40	4·61 ± 0·34
$100\,B/L$	74·2 ± 0·26 (39)	76·1 ± 0·26 (54)	2·44 ± 0·19	2·84 ± 0·18
$100\,H'/L$	72·8 ± 0·41 (30)	75·5 ± 0·29 (51)	3·37 ± 0·29	3·06 ± 0·20
$100\,B/H'$	102·6 ± 0·60 (31)	100·8 ± 0·41 (51)	4·95 ± 0·42	4·34 ± 0·29
$100\,fmb/fml$	83·3 ± 0·72 (29)	—	5·76 ± 0·51	—
[$100\,G'H/GB$]	74·3 ± 0·50 (25)	75·8 ± 0·53 (38)	3·70 ± 0·35	4·83 ± 0·37
[$100\,NB/NH, L$]	49·5 ± 0·58 (29)	48·3 ± 0·48 (42)	4·63 ± 0·41	4·61 ± 0·34
[$100\,O_2/O_1'$]	83·6 ± 0·51 (26)	84·6 ± 0·49 (44)	3·86 ± 0·36	4·81 ± 0·35

* The characters are defined in Appendix I. A symbol in square brackets denotes either that the measurement is one not usually included in the biometric technique, or else that Prof. Derry's method of taking the measurements does not accord with biometric practice.

though there is evidence that the regions represented changed somewhat with time. The series in the first group came from the neighbourhood of Thebes and sites farther south, while those in the second group came from the same region of Upper Egypt and sites farther north. The first group includes all the predynastic series that have been described and some of dynastic date, the latest being of the 18th dynasty: the second group ranges from the 1st dynasty to Roman times, though no series available earlier than the 4th dynasty had come from the region immediately south of the Delta. The Sakkara series described in the present paper extends the range of such material back to the 1st dynasty.

It had been found that the means for all these series are almost constant for most of the metrical characters commonly recorded, but for a few measurements more significant differences are found and these separate the two groups of series. Characters of both kinds are treated in Table 2, which is based on Table XIII in Risdon's paper (1939) on the human remains from Lachish (Palestine). The first six characters are those which make the clearest distinction between the Upper and Lower Egyptian types of series, and they are all breadths or dependent on breadths—the latter being the horizontal circumference and the two indices—of the cranium. The Sakkara series is clearly assigned to the Lower Egyptian group, and if counted as a member of this the range of the mean minimum frontal breadths (B') for the group is slightly extended. The Thebes series is also assigned to the Lower Egyptian group by four of the six characters in question: for U and $100\,B/H'$, however, its means fall within the ranges given for the Upper Egyptian type of series.

Table 2. *Ranges of mean measurements for two groups of series of ancient Egyptian male crania and means for the Sakkara and Thebes series**

Series	Period	B	J	B'	U
Upper Egyptian type	Early predyn.–18th dyn.	131·4–134·3 (10)	123·6–127·5 (8)	90·4–92·8 (4)	500·0–510·4 (4)
Sakkara	1st dyn.	138·7	127·8	96·5	518·8
Thebes	11th dyn.	138·3	127·6	93·6	507·4
Lower Egyptian type	1st dyn.–Roman	135·3–139·3 (9)	127·5–131·3 (8)	93·0–96·2 (5)	510·8–518·7 (5)

Series	Period	$100\,B/L$	$100\,B/H'$	L	H'
Upper Egyptian type	Early predyn.–18th dyn.	71·7–73·7 (10)	98·1–101·1 (10)	182·2–185·2 (10)	132·4–135·9 (10)
Sakkara	1st dyn.	74·2	102·6	186·9	135·4
Thebes	11th dyn.	76·1	100·8	181·8	137·1
Lower Egyptian type	1st dyn.–Roman	73·7–76·0 (9)	102·3–106·4 (9)	181·4–185·8 (9)	130·7–136·0 (9)

* The characters are defined in Appendix I. The numbers in brackets give the numbers of series to which the ranges relate. In the case of these previously described series the smallest number of crania on which any one of the means is based is 16, though this minimum number is about 30 for most of the characters. The numbers on which the Sakkara and Thebes means are based can be seen from Table 1, the only one less than 26 being 14 for the bizygomatic breadth (J) of the Sakkara series.

The last two characters in Table 2, which are the length and height of the cranium, fail to distinguish the two contrasted groups of series. The means for the two new series fall outside the ranges previously given by all the ancient Egyptian material, the Sakkara series giving the greatest L and the Thebes the greatest H'. The evidence of other characters must be taken into account, but so far the comparisons suggest that the two new series are of the Lower Egyptian type, and it is to be expected that they bear a closer resemblance to some of the series assigned to that group than to any other cranial series.

At the same time it may be noted that the Sakkara 1st dynasty and Thebes 11th dynasty populations are clearly differentiated by their mean cranial measurements. There are twenty characters in Table 1 for which means for both series are available. The most significant difference is for L, and it is 6·6 times its probable error, while five other characters—B', U,

prosthion GL, $100\,B/L$ and $100\,H'/L$—also show differences which exceed four times their probable errors.

5. *Comparisons by coefficients of racial likeness.* The method of Karl Pearson's coefficient of racial likeness has been applied extensively to series of ancient Egyptian crania. Risdon (1939) has given comparisons made in that way for twenty-two male series, including three from sites outside Egypt, and the treatment below is almost restricted to comparisons between these and the two new series described in the present paper. The procedure followed in applying the method described in several papers in *Biometrika* was adopted without modification.*

In deriving a classification of a number of cranial, or living, series from the coefficients of racial likeness found between them, it has been shown repeatedly that the most suggestive arrangement is obtained if the closest resemblances of the series, indicated by coefficients below a certain value, are alone taken into account. Risdon has given a diagram (1939, Fig. 3) showing all the reduced coefficients less than 5·0 between the twenty-two series with which he dealt. There are fifty-three of this lowest order among the 231 ($=22\times 21/2$) comparisons. The addition of the two new series to the classification referred to only requires a knowledge of the reduced coefficients less than 5·0 between them and the twenty-two series.

It has been pointed out that inspection of a few mean measurements can indicate whether a comparison of two particular series would almost certainly give a reduced coefficient greater than the limit chosen, or whether it might provide a value less than 5·0. The measurements used for this rapid test are six which are known to be those which show the most significant differences, and the greatest proportions of such differences, in comparisons of the group of series. These are the length, breadth and height of the brain-box and the three indices derived from these chords. For the fifty-three comparisons of the twenty-two

* A 'crude' coefficient is defined by

$$\frac{1}{m}S\left[\frac{n_s n_s'}{n_s+n_s'}\times\frac{(M_s-M_s')^2}{\sigma_s^2}\right]-1\pm 0{\cdot}6745\sqrt{\frac{2}{m}},$$

where M_s is a mean based on n_s crania for the first series, M_s' and n_s' are the corresponding constants for the second series and m characters are compared. The σ's of the long 26th–30th dynasty Egyptian series were used throughout. The crude coefficient may be written

$$\frac{1}{m}S(\alpha)-1\pm 0{\cdot}6745\sqrt{\frac{2}{m}},\quad \text{where}\quad \alpha=\frac{n_s n_s'}{n_s+n_s'}\times\frac{(M_s-M_s')^2}{\sigma_s^2}.$$

Its value is largely determined by the sizes of the two samples that happen to be available, if in fact they do not represent the same population. As many excavated crania are damaged to some extent, in the case of a particular series means for different characters will usually be based on various numbers of specimens (see Table 1). The mean number available for the characters used is denoted by $\bar{n}_s$ in the case of the first series and by $\bar{n}_s'$ in the case of the second series, and these 'sizes' of the samples are usually unequal and may be of very different orders. To obtain, as far as possible, a measure of the absolute divergence of the types compared which does not depend on the numbers of crania available, a 'reduced' coefficient of racial likeness is computed. This is defined to be

$$\frac{100\times 100}{100+100}\times\frac{\bar{n}_s+\bar{n}_s'}{\bar{n}_s\bar{n}_s'}\left[\frac{1}{m}S(\alpha)-1\pm 0{\cdot}6745\sqrt{\frac{2}{m}}\right].$$

A reduced coefficient may be supposed a good approximation to the value which would be obtained if all the means for both series were for 100 individuals instead of for the numbers actually available. If a crude coefficient differs from zero by less than 3·5 times its probable error—a rare occurrence—then it is supposed that there is no evidence of a significant distinction between the two populations represented. In this case there is no need to compute a reduced coefficient. Otherwise, reduced coefficients are found and the classification of a number of series is based on these.

series giving reduced coefficients less than 5·0, the maximum differences for the six characters (in mm. or units of the indices) are:

L	B	H'	$100\,B/L$	$100\,H'/L$	$100\,B/H'$
3·1	3·0	3·5	2·0	2·3	2·8

To avoid the danger of missing comparisons which might be of the order required, in applying the test each of these values was increased arbitrarily by 0·2 giving:

L	B	H'	$100\,B/L$	$100\,H'/L$	$100\,B/H'$
3·3	3·2	3·7	2·2	2·5	3·0

In comparing a new series with the twenty-two it may be supposed that a reduced coefficient of racial likeness greater than 5·0 would almost certainly be found if the difference between the means is greater than the accepted limit in the case of any one or more of the six characters. For such comparisons the coefficients were not calculated. If the differences between the means are less than the limits for all six characters then a reduced coefficient less than 5·0 *might* be found: the coefficients were calculated in all such cases. In this way detailed comparisons were judged to be required between the new Sakkara, 1st dynasty, series, on the one hand, and six of the twenty-two treated by Risdon on the other; and between the new Thebes, 11th dynasty, on the one hand, and only two of the twenty-two series on the other. The previously described series involved in these two sets of comparisons—one series being included in both sets—are:

(i) Deshasheh and Medum, 4th and 5th dynasties (Thomson & MacIver, 1905). The two towns are south of Sakkara and both less than 40 miles from it.

(ii) Gizeh, 26th–30th dynasties (Davin & Pearson, 1924).

(iii) Sedment, 9th dynasty (Woo, 1930).

These three and the new Sakkara series are all from Lower Egypt among the total twenty-four series referred to above. All the other Egyptian sites mentioned are in Middle Egypt and close to Abydos and Thebes.

(iv) Abydos, 18th dynasty (Thomson & MacIver, 1905).

(v) Abydos, 1st dynasty, royal tombs (Morant, 1925).

(vi) Lachish, Palestine (Risdon, 1939). This series represents an Egyptian population. It is assigned to the seventh and eighth centuries B.C., though it is not well dated.

(vii) Tigré district, Abyssinia, modern (Sergi, 1912, means given in Morant, 1925).

(viii) Cretans, modern (von Luschan, 1913, means given in Woo, 1930). This series is not one of the twenty-two dealt with by Risdon. It was included because of its close resemblance to the new 11th dynasty series from Thebes. The test based on a comparison of the means of the six calvarial measurements shows that the only ancient Egyptian series which might give reduced coefficients less than 5·0 with the Cretan series are the Theban 11th dynasty and the Sedment series ((iii) above).

It must be emphasized that a reduced coefficient of racial likeness less than 5·0 represents a very close degree of resemblance. Values of that order have only been found between cranial series which would be expected, on account of their provenance, to represent the same or closely related populations. There is a danger that low reduced coefficients may be misleading owing to the influence on them of extraneous factors, such as inaccuracy in sexing or slight and unappreciated differences between the methods of measurement of two recorders working independently. It is safe to suppose that the two new series are made up entirely of the crania of adult males. In computing coefficients with them care

was taken to restrict a particular comparison to pairs of means based on measurements obtained by following precisely the same technique.

Owing partly to that restriction, the numbers of characters that could be used in computing coefficients with the new series are decidedly smaller than the 31 used ideally for the purpose. For these comparisons the smallest number of characters used is 9 and the largest number is 18.* Risdon (1939, pp. 131–2) has examined the matter experimentally and he concluded that use of a smaller number of characters—the set of 14 he considered being very similar to the sets we were able to use—can usually be expected to give a fairly close approximation to the result which would be obtained from about twice as many

Table 3. *Coefficients of racial likeness between ancient Egyptian, a Palestinian (Lachish) and modern series of male skulls from Abyssinia and Crete**

Series		Crude C.R.L. ± P.E.	Reduced C.R.L.
Sakkara, 1st dyn.	(32·1) with Deshasheh and Medum, 4th and 5th dyn. (46·0)	0·19 ± 0·32 (9)	—
	(32·1) with Abydos, 18th dyn. (49·9)	−0·08 ± 0·32 (9)	—
	(31·6) with Lachish (249·3)	1·04 ± 0·25 (14)	1·85 ± 0·45
	(31·6) with Gizeh, 26th–30th dyn. (885·7)	2·45 ± 0·25 (14)	4·02 ± 0·41
	(31·6) with modern Abyssinian (61·4)	2·43 ± 0·25 (14)	5·82 ± 0·60
	(31·6) with Abydos, 1st dyn. royal tombs (33·6)	1·91 ± 0·25 (14)	5·87 ± 0·77
	(30·3) with Thebes, 11th dyn. (46·7)	4·26 ± 0·22 (18)	11·45 ± 0·59
Thebes, 11th dyn.	(49·2) with Sedment, 9th dyn. (37·9)	−0·43 ± 0·28 (12)	—
	(49·0) with modern Cretans (50·4)	2·01 ± 0·30 (10)	4·04 ± 0·60
	(48·3) with Deshasheh and Medum, 4th and 5th dyn. (46·0)	2·23 ± 0·32 (9)	4·73 ± 0·68
Sedment, 9th dyn.	(37·5) with Deshasheh and Medum, 4th and 5th dyn. (39·9)	1·88 ± 0·25 (14)	4·86 ± 0·65
	(37·7) with modern Cretans (47·9)	2·71 ± 0·25 (15)	6·42 ± 0·59

* The numbers in brackets following the names of the series are the mean numbers of crania for the characters used in computing the coefficients. The numbers in brackets following the crude coefficients are the numbers of characters on which they are based. Woo (1930) gives coefficients with two of the series in the table above, and the values there differ from his because they were recalculated omitting the term $1/m$, which was discarded after 1930. The standard deviations of the long E series of 26th–30th dynasty crania from Gizeh (Davin & Pearson, 1924) were used in computing all the coefficients in the table.

characters. Occasionally, however, use of a smaller number of characters may suggest a rather misleading conclusion, and it will tend to indicate a rather wider separation of the series than that which would be found if all 31 characters could be used. With these reservations in mind the coefficients with the new series may be accepted as the best approximations it is possible to obtain in the circumstances.

All the coefficients of racial likeness found with the Sakkara 1st dynasty and Thebes 11th dynasty series are given in Table 3. Fig. 1 is a reproduction of part of a diagram given by Risdon (1939, p. 137) for the twenty-two ancient Egyptian and related series with which he dealt, with the addition of the two series described in the present paper and that of modern Cretans. The Sakkara series is seen to be an unexceptional member of the 'Lower Egyptian' constellation, having two insignificant coefficients and other close connexions

* The characters common to all the comparisons with the new series are L, B, H', LB, J, NB, $100\,B/L$, $100\,H'/L$ and $100\,B/H'$. Others used in some cases are B', U, S, fml, fmb and $100\,fmb/fml$, and for the coefficient between the two new series only $G'H$, NH, L, O_1', O_2, $100\,G'H/GB$, $100\,NB/NH$, L, $100\,O_2/O_1'$.

with members of that group. On the other hand, the 11th dynasty soldiers from Thebes clearly represent an Egyptian population of an aberrant type. The direct comparison fails to distinguish this from that of the 9th dynasty series from Sedment. Woo (1930) had found that the latter stands apart from all the other Egyptian series, and he pointed out that the Sedment bears a closer resemblance to a series of modern Cretans (von Luschan, 1913) than to most of the ancient Egyptian series. The Thebes 11th dynasty series has a reduced coefficient less than 5·0 with the Cretan, though the latter has no other coefficient of this order with any of the other Egyptian series.

6. *The racial history of ancient Egyptian populations.* The new evidence makes rather more precise the racial classification of ancient Egyptian populations given in earlier craniological papers in *Biometrika*. The 1st dynasty series of crania from Sakkara is the earliest in date that has been described representing the region immediately south of the Delta. It is an unexceptional representative of that group which must have prevailed

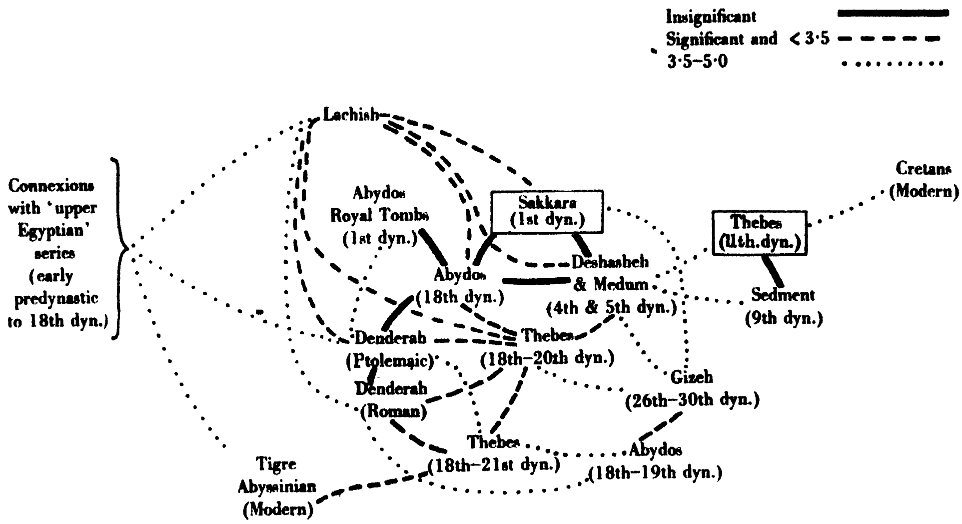


Fig. 1. Reduced coefficients of racial likeness between the two new and other ancient Egyptian and related series of male crania.

in the region—with only slight local and secular variants—from the 1st to the 30th dynasty and probably in both earlier and later periods as well. Such populations are said to be of 'Lower Egyptian' type.

In the region to the south, round Thebes and Abydos, the population was of a second racial type from the earliest predynastic (Badari) epoch for which there is any adequate craniological evidence. This is called the 'Upper Egyptian', though it would be better to call it Southern Egyptian. The population became modified slowly down to some time about the 18th dynasty. The change was such that the 'Upper Egyptian' type of population came to bear a closer and closer resemblance to the 'Lower Egyptian', though the two groups remained clearly distinct. About the 18th dynasty there must have been a fairly rapid, if not abrupt, change in the racial composition of the population of the Thebes and Abydos region. Nearly all the series from there, of that and later dates, are not of 'Upper' but of 'Lower Egyptian' type. They diverge slightly from the populations of the region immediately south of the Delta, however, in the direction of the 'Upper Egyptian' type. Six of these series—viz. those from Abydos, Thebes and Denderah of

dates ranging from the 18th dynasty to Roman times—are shown in Fig. 1. The 1st dynasty series from royal tombs at Abydos, also shown there, is an exception on account of its date. The obvious explanation of its peculiar position is that it represents an intrusive and more or less isolated community which was derived from the other centre of population to the north.

This accounts for twenty-two of the twenty-four series of crania considered. The classification of these does not seem to necessitate reference to any non-Egyptian peoples. This is not so, however, in the case of the remaining two series, viz. the new one of 11th dynasty soldiers from Thebes and the 9th dynasty series from Sedment (Deltaic region). These two might represent the same population as far as can be seen from the direct comparison, and both stand apart from the 'Lower Egyptian' constellation of series (see Fig. 1). The fact that the 11th dynasty series from Thebes has a close resemblance to one of Cretans, which is of modern date, suggests that the two aberrant communities in question may have been derived from the crossing of ancient Egyptians with people from some European or Asiatic source.

The mean basio-bregmatic heights (H'), cephalic indices and height-length indices are higher for the 11th dynasty Thebes and Sedment series than for any other of the series considered. The types, defined by average measurements, of these two thus diverge from that prevailing in ancient Egypt in the direction of the 'Armenoid' type. Elliot Smith (1911 and elsewhere) supposed that intrusive 'Armenoid' aliens played a considerable part in modifying the population of the country and that 'long before the time of the New Empire, Egypt was permeated from one end to the other with this foreign element'.

Our interpretation of the evidence fails entirely to support this hypothesis. There is no need to suppose that any people foreign to the country played a substantial part in modifying its population from predynastic to Roman times. The communities represented by the 11th dynasty Thebes and Sedment series may possibly have been derived from the crossing of Egyptian and 'Armenoid' people, but they stand apart. The remarkable point is not that two out of twenty-four populations should be peculiar in that way, but that the remaining twenty-two show interrelationships which do not suggest any admixture with alien stock. They can readily be explained on the supposition that there was a steady transference of population from the Deltaic region to the region of Thebes and Abydos, where the population was originally of a somewhat different type, from early predynastic times to the 18th dynasty. About that time the movement must have been accelerated, and thereafter the populations of the two centres were almost indistinguishable in racial type. The racial history of ancient Egypt was of a simple kind.

7. *Summary and conclusions.* This paper deals with forty-four male crania of 1st dynasty date from Sakkara and with fifty-five crania of 11th dynasty soldiers from Thebes. Individual measurements taken by Prof. D. E. Derry are given in appended tables. Judging from the rather small samples, the two populations represented exhibited the same order of variation, while both were rather less mixed in racial composition than the population of Giza from the 26th–30th dynasties. Mean measurements clearly differentiate the two new series from one another. Judging from characters considered singly, both series bear a close resemblance to some other ancient Egyptian series, and both are of 'Lower' rather than 'Upper Egyptian' type. Comparisons are made by the method of the coefficient of racial likeness, though decidedly fewer characters than the standard set of thirty-one used when possible are available for the purpose. The resulting relationships are shown in Fig. 1.

Appendix II. Individual measurements of the First Dynasty male crania from Sakkara*

Grave no.	L	B	B'	H'	[Aur. M.]	LB	U	[G'H]	OB	J	[NH, L]	NB	[O_1']	[O_2]	[Prosthion GL]	fml	fmb	$100\frac{B}{L}$	$100\frac{H'}{L}$	$100\frac{B}{H'}$	$\left[100\frac{G'H}{GB}\right]$	$\left[100\frac{NB}{NH', L}\right]$	$\left[100\frac{O_2}{O_1'}\right]$	$100\frac{fmb}{fml}$	Remarks
3	194	135·5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	69·8	—	—	—	—	—	—	—
4	183·5	136	98·5	124	107	100	513	68·5	93	—	52·5	24·5	41	35·5	96?	40	31·5	74·1	67·6	109·7	73·7	46·7	86·6	78·7	About 18 years. Metopic
7	184	137·5	92·5	131·5	113	100	513·5	72	103	120	52·5	27	37·5	30·5	96	38·5	30·5	74·7	71·5	104·6	69·9	50·4	81·3	79·2	Skull female? but pelvis definitely male
8	192	137	—	—	121	—	—	72?	96	—	50·5	22·5	39·5	31·5	—	—	—	71·3	—	—	75·0?	44·6	79·7	—	—
10	191	147	97·5	144	125	98	540	66	88	—	45	23·5	38	30·5	101	34	30·5	77·0	75·4	102·1	75·0	52·2	80·3	89·9	Root of nose depressed. Slight hydrocephaly?
14	180·5	141·5	97	135·5	115	100	517·5	70·5	99	132	52	23·5	41	32	96	35	29	75·9	72·7	104·4	71·2	45·2	78·0	82·9	—
16	180·5	139	—	140	116·5	101	503	73·5	97	—	54·5	29	35·5	33·5	101	36	31·5	77·0	77·6	99·3	75·8	53·2	94·4	87·5	About 16 years
17	189	139·5	100·5	143	—	106·5?	—	—	—	—	—	—	—	—	—	38·5	31·5	73·8	75·7	97·6	—	—	—	81·8	—
18	179	134	90	—	—	—	—	—	—	—	—	—	—	—	—	—	—	74·9	—	—	—	—	—	—	—
21	—	147	—	134·5	—	—	—	—	—	—	—	27·5	—	—	—	37	29	—	—	109·3	—	—	—	78·4	—
23	—	141·5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
27	—	138·5	—	127	—	—	—	—	—	—	—	—	—	—	—	36·5	34	—	—	109·1	—	—	—	93·1	—
30	181·5	139	94·5	132	117	98	509	63	89·5	—	44·5	24	40	32	101	36	32	76·6	72·7	105·3	70·4	53·9	80·0	88·9	—
32	189·5	136·5	93	130·5	116	95	525	76	100·5	129	56	28·5	40	34	95·5	40	31	72·1	68·9	104·6	75·6	50·4	85·0	77·5	—
38	193	140	99	134	—	107·5	—	77·5	—	—	56·5	26·5	39·5	32·5	101	34	31·5	72·3	69·4	104·5	—	46·9	82·3	92·6	—
49	184·5	134·5	98	136	113	94	516	72·5	101	127	48·5	26	38·5	32	98	39	31	72·9	73·7	98·9	71·8	53·7	83·1	79·5	—
50	180·5	138	94	143	114·5	102	510·5	69	92·5	124	48	25	35	31·5	96?	37·5	31·5	76·5	79·2	96·5	74·0	52·1	90·0	84·1	About 19 years. Skull female? but pelvis definitely male
56	186·5	—	101	128	103·5	107	—	64	93·5	—	46	25	38	31	96·5	31	29	—	68·6	—	68·4	53·2	81·6	93·5	Distorted by grave pressure
58	189	137	93·5	145	119	107	520·5	70	100·5	—	48·5	24·5	40·5	32	105	39	32	72·5	76·7	94·5	69·7	50·5	79·0	82·1	—
59	189	143	101·5	137	120	99	527	75·5	—	—	55	27	—	—	90?	37·5	—	75·7	72·5	104·4	—	49·1	—	—	—
62	182	141	97·5	132·5	107	100·5	518	67?	90	126	46	25	40·5	32	95?	35	31	77·5	72·8	106·4	74·4?	53·2	79·0	88·6	About 18 years
65	189	138·5	95·5	—	—	—	—	78·5	97	—	57·5	25·5	—	—	—	—	—	73·3	—	—	80·9	44·3	—	—	—
72	178	139	91·5	127·5	109	100	501	75·5	100·5	—	56·5	23	38	33	101·5	36	29	78·1	71·6	109·0	75·1	40·7	86·8	80·6	—
77	195	138·5	103?	—	—	—	—	—	—	—	—	—	—	—	—	—	—	71·2	—	—	—	—	—	—	Metopic
82	195	138·5	96	141	121·5	111	535	77·5	106	131	57	22·5	40	34	105	36	26·5	71·0	72·3	98·2	73·1	39·5	85·0	73·6	—
87	181	139·5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	77·1	—	—	—	—	—	—	About 17 years
88	188	134·5	97·5	139	—	—	510	—	—	—	—	—	—	—	—	—	—	71·5	73·9	96·8	—	—	—	—	—
90	193·5	138	91·5	131	—	104	524	—	—	—	—	—	—	—	—	39	28	71·3	67·7	105·3	—	—	—	71·8	—
91	192·5	147·5	96	133	—	101·5	535	—	—	—	—	—	—	—	—	39	35	76·6	69·1	110·9	—	—	—	89·7	—
92	178·5	132·5	94	—	—	—	507	—	—	—	—	—	—	—	—	—	—	74·2	—	—	—	—	—	—	—
98	183	134·5	93·5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	73·5	—	—	—	—	—	—	About 18 years
99	187·5	139·5	103	145	118	107	525	72	96	135	48·5	25·5	41	32·5	100	32·5	29·5	74·4	77·3	96·2	75·0	52·6	79·3	90·8	—
121	186	137	95·5	136	116	106	518	72·5	104	130	50·5	25	38	31·5	105	36	31·5	73·7	73·1	100·7	69·7	49·5	82·9	87·5	—
122	179·5	143·5	99	131	113	102·5	511·5	67	91·5	126·5	48·5	26·5	38·5	31·5	99	37·5	30·5	79·9	73·0	109·5	73·2	54·6	81·8	81·4	About 18 years
124	180	132·5	97	139·5	119	101·5	501?	70·5	93·5	122·5	50	26	38·5	33	100	35	28·5	73·6	77·5	95·0	75·4	52·0	85·8	81·5	About 20 years
128	185·5	139	88·5	133	113	97·5	511	74	98·5	122·5?	54·5	23	—	—	97·5	37	29·5	74·9	71·7	104·5	75·1	42·2	—	79·7	About 20 years
141	184?	134	99	141	116	101·5	515?	68·5	—	—	49	28	36	32	95	39·5	—	72·8?	76·6?	95·0	—	57·1	88·9	—	Metopic. Distorted
145	195	130	100	132	109	109·5	537	74	—	—	—	—	37·5	31	104·5	38	29·5	71·3	67·7	105·3	—	—	82·7	77·5	—
151	197	137	—	—	—	—	—	70·5	—	·	43·5	25	—	—	—	—	—	69·5	—	—	—	57·5	—	—	—
175	189	138·5	93	129	107	106·5	516	77	96·5	132	52·5	27	41	33	106·5	37	30·5	73·3	68·3	107·4	80·8	51·4	80·5	82·4	Old
180	184	142	96	142	120	102·5	520	77·5	94	125·5	56	26	38	33·5	100	33·5	28	77·2	77·2	100·0	82·4	46·4	88·2	83·8	—
184	194	146·5	105	—	119	—	553	79	97·5	—	53	25	43	37·5	—	—	—	75·5	—	—	81·0	47·2	87·2	—	—
190	185	130	96·5	134·5	112·5	112·5	—	67	94	—	51	24	38	32·5	109	37	29	70·3	72·7	96·7	71·3	47·1	85·5	78·4	About 17 years. Negroid
226	188·5	142·5	93·5	—	—	—	512?	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

* The measurements are defined in Appendix I. A symbol above in square brackets denotes either that the measurement is one not usually included in the biometric technique, or else that Prof. Derry's method of taking the measurement does not accord with biometric practice.

APPENDIX III. INDIVIDUAL MEASUREMENTS OF THE CRANIA OF ELEVENTH DYNASTY SOLDIERS FROM THEBES*

Grave no.†	L	B	B'	H'	LB	[Broca's Q']	S	S_1	S_2	S_3	U	[$G'H$]	GB	J	[NH, L]	NB	[O_1']	[O_2]	[Prosthion GL]	$100\frac{B}{L}$	$100\frac{H'}{L}$	$100\frac{B}{H'}$	$\left[100\frac{G'H}{GB}\right]$	$\left[100\frac{NB}{NH, L}\right]$	$\left[100\frac{O_2}{O_1'}\right]$	Remarks
2	180·5	138	90	138	101	296	366·5	125·5	129	112	502	72·5	102·5	128	51·5	24	38·5	31	101	76·5	76·5	100·0	70·7	46·6	80·5	—
4	171·5	140	93	136·5	95·5	310	356	123	135	98	495	68·5	88·5	120·5	49	23	38	32·5	88·5	81·6	79·6	102·6	77·4	46·9	85·5	—
6	181·5	132·5	90	134·5	102·5	283	356	120	121	115	498	78	—	—	57	22·5	39	37	95	73·0	74·1	98·5	—	39·5	94·9	—
9	179·5	136	100	138·5	108	301	354	116	129	109	497	69	100·5	134	50	30	42	32	108·5	75·8	77·2	98·2	68·7	60·0	76·2	Metopic
10	181·5	135·5	93·5	139	102·5	295	366	116	123	127	510	72·5	92·5	122	53	23	38	33	93·5	74·7	76·6	97·5	78·4	43·4	86·8	Metopic
14	187	133·5	95·5	139·5	100·5	310	389	137	137	115	—	75	102	133	50	25·5	38·5	35	102	71·4	74·6	95·7	73·5	51·0	90·9	—
18	184	139·5	90	136·5	100	294	368	128	121	119	506	74	96·5	130	53·5	25	41	37?	95·5	75·8	74·2	102·2	76·7	46·7	90·2?	—
21	198·5	139	98	144	106·5	310	401	131	130	140	535	70	98·5	130	50·5	25	42·5	34	99·5	70·0	72·5	96·5	71·1	49·5	80·0	—
23	185·5	138·5?	86	133	103	293	369	123	130	116	—	—	—	—	—	—	—	—	—	74·7?	71·7	104·1?	—	—	—	—
25	177	138	94	140	98	303	371	132	134	105	508	77	94·5?	127·5	53	24·5	37	33·5	89	78·0	79·1	98·6	81·5?	46·2	90·5	—
26	182	128	98	144	107·5	295	374	124	144	106	498	79·5	96	132·5	55	25	41·5	35	100·5	70·9	79·1	88·9	82·8	45·5	84·3	Left parietal fractured
28	175	133	88	141	99	291	359	119	127	113	487	69·5	93	128·5	49	26	41	30	93	76·0	80·6	94·3	74·7	53·1	73·2	—
29	181·5	138·5	93	131	99·5	297	365	125	125	115	508	75·5	—	—	55·5	26	39	37	98	76·3	72·2	105·7	—	46·8	94·9	—
44	180	136	100	127·5	100·5	290	355	119	119	117	503	69	96	121	52	25	39	29·5	95·5	75·5	70·8	106·7	71·9	48·1	75·6	—
45	178	130	95	132	96·5	280	366	128	130	108	495	69	90	125	51	24	38	29	91	73·0	74·1	98·5	76·7	47·1	76·3	—
46	179·5	132	87	135	93	283	375·5	128	129	118·5	488	64·5	96·5	126	48·5	24	38	32	94·5	73·5	75·2	97·8	66·8	49·5	84·2	—
47	178·5	137	94	134	96·5	293	359	121	119	119	501	71	99	133	52	25	39	32	91	76·7	75·1	102·2	71·7	48·1	82·1	—
48	185	143	93·5	129	101	310	364	122	120	122	524	70·5	93	—	50	30	38	33·5	96	77·3	69·7	110·8	75·8	60·0	88·2	Metopic
49	175	141	86·5	143	96·5	305	367	127	119	121	495	71·5	81	119	50?	20	39	33·5	92·5	80·6	81·7	98·6	88·3	40·0?	85·9	—
50	184	142·5	94·5	137	98	308	383	123	136	124	520	68·5	97	—	51·5	27	39	32·5	92	77·4	74·5	104·0	70·6	52·4	83·3	—
51	196	141	100	137	101	312	399	130	137	132	541	75	99	131	50	25·5	40·5	32	104	71·9	69·9	102·9	75·7	51·0	79·0	—
52	183	142·5	95	139	102	313	368	130	130	108	512·5	69·5	—	—	50	24	41	34·5?	97	77·9	75·9	102·5	—	48·0	84·2?	Fractures in parietal, frontal and orbital regions
53	178	141·5	103	144	105	315	359	127	129	103	512	78	97	129·5	57·5	24·5	41	36·5	91·5	79·5	80·9	98·3	80·4	42·6	89·0	—
54	178	130	91	139·5	103·5	291	370	126	138	106	500	74	95	124	53	24·5	37·5	34·5	95·5	73·0	78·4	93·2	77·9	46·2	92·0	—
55	185	140	97	137·5	95	296	376	125	123	128	520	70	94	127·5	53	23	39	35	89·5	75·5	74·1	101·8	74·5	43·4	89·7	Metopic
56	188	144	94·5	142	103	307	381	132	133	116	530	74	94	—	54	24·5	40	33	98·5	76·6	75·5	101·4	78·7	45·4	82·5	Metopic
57	180	145	97	—	—	304	371	124	133	114	519	75·5	91·5	127·5	53	24·5	37·5	32·5	—	80·5	—	—	82·5	46·2	86·7	Metopic
58	180	140	97	139·5	100	305	376	127	133	116	506	69·5	100?	—	49	—	—	—	103·5	77·8	77·5	100·4	69·5?	—	—	Metopic
59	191	139	101·5	147·5	105	312	401	142	133	126	523	77·5	97·5	138	57	27	40	37·5	103·5	72·8	77·2	94·2	79·5	47·4	93·8	—
60	180	148·5	99	138	97·5	315	382	127	135	120	527	74·5	93	129	49	26	38·5	33·5	100	82·5	76·7	107·6	80·1	53·1	87·0	Metopic
62	180	139	100	133	101·5	—	369	123	135	111	—	76	98	128	53	24	40·5	35	96	77·6	73·9	104·5	77·5	45·3	86·4	Metopic
63	177·5?	140?	100?	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	78·9?	—	—	—	—	—	Metopic
65	181	139·5	87	133	97·5	297	367	123	128	116	502	68	—	—	49	25	37·5	30	95·5	77·1	73·5	104·9	—	51·0	80·0	—
66	190	144	89	139·5	98·5	310	393	135	140	118	545	68	105	—	52	27	39	32·5	92·5	75·8	73·4	103·2	64·8	51·9	83·3	—
67	200	149	93	138·5	96	316	400	132	139	129	533	75	99	130	54	22	40	32·5	97·5	74·5	69·3	107·6	75·7	40·7	81·3	—
68	190	135	90	139·5	103	301	388	128	136	124	520	69·5	94	129	48	24·5	38·5	32·5	104	71·1	73·4	96·8	73·9	51·0	84·4	—
69	175	133	86	138	103	292	360	121	135	104	487	69·5	88	124	48·5	24	36·5	29	100·5	76·0	78·9	96·4	79·0	49·5	79·5	—
70	184	134	93	133·5	100	296	372	130	130	112	507	65·5	91	124·5	50	30	38	31	97·5	72·8	72·5	100·4	72·0	60·0	81·6	—
71	175	142	102	135	101·5	310	359	122	134	103	506	72	97	132	52	25·5	41	36	96	81·1	77·1	105·2	74·2	49·0	87·8	—
72	173	134	97	135	104	296	350	123	122	105	487	74·5	93	122·5	54·5	23·5	38	33·5	96	77·5	78·0	99·3	80·1	43·1	88·5	—
73	182	136	88·5	131	97	293	367	123	124	120	501	71	—	—	54	—	38·5	32	94·5	74·7	72·0	103·8	—	—	83·1	—
74	176·5	135·5	85	142	100	303	363	120	135	108	491	72	92	122·5	49	24	37	29	89·5	76·8	80·5	95·4	78·3	49·0	78·4	—
75	178	135	103	133	106	301	356	133	122	101	500	78	99·5	128	57	25·5	42	33·5	96	75·8	74·7	101·5	78·4	44·7	79·8	—
76	188	143	90	135	106·5	—	369	131	122	116	511	67	—	—	52	26	42	33	102	76·1	71·8	105·9	—	50·0	78·6	—
77	174·5	134	91·5	138	102	297	351	123	125	103	484	71·5	99	127·5	50·5	24·5	37·5	31·5	98·5	76·8	79·1	97·1	72·2	48·5	84·3	Metopic
78	183·5	144	93	132	100	300	374	123	128	123	521	—	—	—	—	—	—	—	—	78·5	71·9	109·1	—	—	—	Metopic
79	183	133	99·5	134	—	289	361	125	128	108	508	—	—	—	—	—	—	—	—	72·7	73·2	99·3	—	—	—	Metopic
80	—	—	97·5	—	—	—	—	—	—	—	—	73·5	94	—	51·5	26	37·5	32	—	—	—	—	78·2	50·5	85·4	Metopic
81	182·5	140	94	143	98	295	378	125	125·5	128	508	67?	—	—	50	—	38	35	92·5	76·7	78·3	97·9	—	—	92·1	—
82	181·5	142	86	137	—	305	378	128	130	120	505	—	—	—	—	—	—	—	—	78·2	75·5	103·6	—	—	—	—
83	178·5	135	93·5	136?	—	294	—	122	126	—	495	—	—	—	—	—	—	—	—	75·6	76·2?	99·3?	—	—	—	—
84	183·5	139	83	—	—	—	—	122	127	—	502	—	—	—	—	—	—	—	—	75·7	—	—	—	—	—	—
85	174·5	140	93	137	102	303	362	123	130	109	501	—	—	—	—	—	—	—	—	80·2	78·5	102·2	—	—	—	—
86	177	138	89	138	—	298	361	124	125	112	500	—	—	—	—	—	—	—	—	78·0	78·0	100·0	—	—	—	—
87	183?	139	94	135	—	—	—	—	130	115	514	—	—	—	—	—	—	—	—	75·9	73·8	103·0	—	—	—	—

* The measurements are defined in Appendix I. A symbol above in square brackets denotes either that the measurement is one not usually included in the biometric technique, or else that Prof. Derry's method of taking the measurement does not accord with biometric practice.

† Only fifty-nine bodies were counted. The skulls were numbered separately when they could not be definitely associated with particular skeletons and hence the numbers used exceed fifty-nine.

The Sakkara 1st dynasty series, which is the earliest from the region immediately south of the Delta, is an unexceptional member of the 'Lower Egyptian' constellation, and it can be supposed to typify the population of Northern Egypt at the time. The 11th dynasty series of soldiers from Thebes is linked to the same group, but it diverges from it. The type is indistinguishable from that of a 9th dynasty series, from Sedment. The former also has a link with the type of a series of modern Cretans. The two aberrant communities of Thebes and Sedment must be supposed to have been derived from the crossing of ancient Egyptians with people from some European or Asiatic source. Our knowledge of the racial history of ancient Egypt derived from craniological evidence is reviewed.

REFERENCES

DAVIN, A. G. & PEARSON, K. (1924). On the biometric constants of the human skull. *Biometrika*, **16**, 328–63.

DUCKWORTH, W. L. H. (1913). *International Agreements for the Unification (a) of Craniometric and Cephalometric Measurements, (b) of Anthropometric Measurements to be made on the Living Subject.* Camb. Univ. Press.

LUSCHAN, F. VON (1913). Beiträge zur Anthropologie von Kreta. *Z. Ethn.* **14**, 307–93.

MACRAMALLAH, R. (1940). *Un Cimetière Archaïque de la Classe Moyenne à Saqquara.* Imp. Nat. le Caire.

MORANT, G. M. (1925). A study of Egyptian craniology from prehistoric to Roman times. *Biometrika*, **17**, 1–52.

MORANT, G. M (1935). A study of predynastic Egyptian skulls from Badari based on measurements taken by Miss B. N. Stoessiger and Prof. D. E. Derry. *Biometrika*, **27**, 293–309.

RISDON, D. L. (1939). A study of the cranial and other human remains from Palestine excavated at Tell Duweir (Lachish) by the Wellcome-Marston archaeological research expedition. *Biometrika*, **31**, 99–166.

SERGI, S. (1912). *Crania Habessinica. Contributo all' Antropologia dell' Africa Orientale.* Rome: Loescher.

SMITH, G. ELLIOT (1911). *The Ancient Egyptians and their Influence upon the Civilization of Europe.* Harper.

STOESSIGER, B. N. (1927). A study of the Badarian crania recently excavated by the British School of Archaeology in Egypt. *Biometrika*, **19**, 110–50.

THOMSON, A. & RANDALL-MACIVER, D. (1905). *The Ancient Races of the Thebaid.* Oxford Univ. Press.

WINLOCK, H. E. (1928). The Egyptian Expedition, 1925–27. *Bull. Met. Mus. Art, New York.*

WOO, T. L. (1930). A study of seventy-one ninth dynasty Egyptian skulls from Sedment. *Biometrika*, **22**, 65–93.

APPENDIX I. DEFINITIONS OF MEASUREMENTS

Individual measurements of the two series of crania are given in Appendices II–III. The contractions used there and in tables in the text to denote characters are:

L = maximum glabella-occipital length. B = maximum horizontal breadth. B' = minimum frontal breadth. H' = basio-bregmatic height. *Aur. ht.* = 'vertical height from line joining highest points of external auditory meatuses'. LB = basion to nasion. U = maximum horizontal circumference above the superciliary ridges. S_1 = arc nasion to bregma. S_2 = arc bregma to lambda. S_3 = arc lambda to opisthion. S = total sagittal arc from nasion to opisthion. *Broca's* Q' = transverse arc from 'the most prominent point on the posterior root of the left zygoma, exactly above the auditory aperture', to the same point on the right passing through the bregma. fml = basion to opisthion. fmb = maximum breadth of foramen magnum. $G'H$ = nasion to alveolar point. GB = facial breadth between lowest points on zygomatico-maxillary sutures. J = maximum breadth between zygomatic arches. NH, L = nasal height from nasion to point furthest removed from it on the margin of the left pyriform aperture. NB = maximum breadth of the pyriform aperture. O_1' = breadth of right orbit from the dacryon. O_2 = maximum height of right orbit. *Prosthion* GL = basion to prosthion.

THE GENERALIZATION OF 'STUDENT'S' PROBLEM WHEN SEVERAL DIFFERENT POPULATION VARIANCES ARE INVOLVED

By B. L. WELCH, B.A., Ph.D.

1. *Introduction and summary.* Let η be a population parameter which is estimated by an observed quantity y, normally distributed with variance σ_y^2. Let $\sigma_y^2 = \sum_{i=1}^{k} \lambda_i \sigma_i^2$, where the λ_i are known positive numbers and the σ_i^2 are unknown variances. Suppose that the observed data provide estimates s_i^2 of these variances, based on f_i degrees of freedom, respectively, so that the sampling distribution of s_i^2 is

$$p(s_i^2)\, ds_i^2 = \frac{1}{\Gamma(\frac{1}{2}f_i)}\left(\frac{f_i s_i^2}{2\sigma_i^2}\right)^{\frac{1}{2}f_i-1} \exp\left[-\frac{1}{2}\frac{f_i s_i^2}{\sigma_i^2}\right] d\left(\frac{f_i s_i^2}{2\sigma_i^2}\right), \tag{1}$$

and that these estimates are distributed independently of each other and of y.

A very simple particular case of this set-up occurs when we have samples of n_1 and n_2, respectively, from two normal populations with true means α_1 and α_2 and standard deviations σ_1 and σ_2. If η is the true difference $(\alpha_1 - \alpha_2)$ between the means, the estimated difference is $y = (\bar{x}_1 - \bar{x}_2)$. The variance of the estimate is $\sigma_y^2 = (\lambda_1\sigma_1^2 + \lambda_2\sigma_2^2)$, where $\lambda_1 = 1/n_1$ and $\lambda_2 = 1/n_2$. The estimated values of σ_1^2 and σ_2^2 are $s_1^2 = \Sigma_1/f_1$ and $s_2^2 = \Sigma_2/f_2$, where Σ_1 and Σ_2 are the respective sums of squares of observations from the individual sample means and $f_1 = (n_1 - 1)$ and $f_2 = (n_2 - 1)$. These s^2 are distributed in the form (1) and the postulated conditions of independence hold.

Another particular case, again with $k = 2$, arises when we wish to compare two regression coefficients, fitted to independent sets of data, without making the assumption that the population residual variance about the true regression line is the same for both sets.

The present paper is written mainly with these practical applications of the case $k = 2$ in mind, but the results are expressed generally for any k, since no further analytical difficulties are involved. It will be shown how probability statements about y, considered as an estimate of η, may be made similar in character to those which W. S. Gosset derived for the mean of a single sample of n observations ('Student', 1908). We shall, in effect, seek a quantity h, calculable from the observations, with the property that the chance of the difference $(y - \eta)$ falling short of h is a given probability P. It is clear that h must be a function of the individual variances s_i^2 and of P. If the abbreviation Pr. is used to mean 'the probability of the relation in the bracket following', our problem is to satisfy the equation

$$\Pr.\,[(y-\eta) < h(s_1^2, s_2^2, \ldots, s_k^2, P)] = P. \tag{2}$$

In Gosset's case the solution was, of course, simply

$$\Pr.\,[(\bar{x} - \alpha) < t_P s/\sqrt{n}] = P, \tag{3}$$

where t_P is the value, corresponding to the probability level P, in the 'Student' t-distribution with $f = (n-1)$ degrees of freedom.

In the next section the mathematical derivation of the exact solution of (2) is given. This is then followed by some consideration of its expression in numerical terms. First, a series solution in powers of $1/f_i$ is developed, which may be used for calculating tables. Then some comparisons are made with a non-series approximate solution which is based on a particular way of regarding the distribution of a quantity of the general form $z = (\Sigma a_i \chi_i^2)$.

Some brief discussion is then added which may serve to place the present contribution in its proper relationship to other papers which have been written on this topic.

Finally, it is shown how the inequality (2) may be adapted to provide an interval estimate for η.

2. *Mathematical derivation of solution.* Let $j(s_1^2, s_2^2, \ldots, s_k^2, P)$ denote the probability that $(y-\eta)$ is less than $h(s_1^2, s_2^2, \ldots, s_k^2, P)$, *given* s_i^2 $(i = 1, 2, \ldots, k)$. Then, since y is distributed quite independently of the estimated variances, we have

$$j(s_1^2, s_2^2, \ldots, s_k^2, P) = \int_{u=-\infty}^{\frac{h(s_1^2, s_2^2, \ldots, s_k^2, P)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}u^2}du = I\left\{\frac{h(s_1^2, s_2^2, \ldots, s_k^2, P)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right\}, \tag{4}$$

where I is used to denote the normal probability integral. The condition of equation (2) is then simply that, if $j(s_1^2, s_2^2, \ldots, s_k^2, P)$ is averaged over the probability distributions of s_i^2 as given by (1), the result will equal P. Thus

$$\int_{s^2} \cdots \int j(s_1^2, s_2^2, \ldots, s_k^2, P) \prod_i p(s_i^2)\, ds_i^2 = P. \tag{5}$$

Now we may expand $j(s_1^2, s_2^2, \ldots, s_k^2, P)$ about an origin $(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2)$ in a Taylor expansion. Thus

$$j(s_1^2, s_2^2, \ldots, s_k^2, P) = \exp\left[\sum_i (s_i^2 - \sigma_i^2)\,\partial_i\right] j(w_1, w_2, \ldots, w_k, P), \tag{6}$$

it being understood that the exponential is to be expanded in a power series in ∂_i and that ∂_i is to be interpreted so that

$$\partial_i^r j(w_1, w_2, \ldots, w_k, P) = \left[\frac{\partial^r}{\partial w_i^r} j(w_1, w_2, \ldots, w_k, P)\right]_{w_i = \sigma_i^2}. \tag{7}$$

On making the substitution of (6) into (5) our result may be written

$$\Theta j(w_1, w_2, \ldots, w_k, P) = P, \tag{8}$$

where
$$\Theta = \prod_i \int \exp\left[(s_i^2 - \sigma_i^2)\partial_i\right] p(s_i^2)\, ds_i^2. \tag{9}$$

Now, substituting into (9) from (1), the integral comes out in simple form, i.e.

$$\begin{aligned}
\Theta &= \prod_i \left\{1 - \frac{2\sigma_i^2\partial_i}{f_i}\right\}^{-\frac{1}{2}f_i} \exp\left[-\sigma_i^2\partial_i\right] \\
&= \exp\left\{-\Sigma\sigma_i^2\partial_i - \tfrac{1}{2}\Sigma f_i \log\left(1 - \frac{2\sigma_i^2\partial_i}{f_i}\right)\right\} \\
&= \exp\left\{\Sigma\frac{\sigma_i^4\partial_i^2}{f_i} + \frac{4}{3}\Sigma\frac{\sigma_i^6\partial_i^3}{f_i^2} + 2\Sigma\frac{\sigma_i^8\partial_i^4}{f_i^3} + \text{etc.}\right\} \\
&= 1 + \Sigma\frac{\sigma_i^4\partial_i^2}{f_i} + \left\{\frac{4}{3}\Sigma\frac{\sigma_i^6\partial_i^3}{f_i^2} + \frac{1}{2}\left(\Sigma\frac{\sigma_i^4\partial_i^2}{f_i}\right)^2\right\} + \text{etc.}
\end{aligned} \tag{10}$$

Substituting (4) into (8) we have finally

$$\Theta I\left\{\frac{h(w_1, w_2, \ldots, w_k, P)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right\} = P. \tag{11}$$

This, in a very condensed form, is the solution to our problem.* The operator Θ constitutes a direction to carry out the partial differentiations indicated by (10). w_j must then be equated to σ_j^2. The solution of the resulting equation will give $h(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2, P)$ and therefore the required $h(s_1^2, s_2^2, \ldots, s_k^2, P)$.

3. *The development of the series solution.* It will be convenient to write $h(w)$ for $h(w_1, w_2, \ldots, w_k, P)$ and ξ for the normal deviate such that $I(\xi) = P$. We may then expand $I\left\{\frac{h(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right\}$ in a Taylor series about ξ as origin. Thus

$$I\left\{\frac{h(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right\} = \exp\left[\left\{\frac{h(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}} - \xi\right\} D\right] I(v), \tag{12}$$

it being understood that the exponential is to be expanded in powers of D, and that these powers are to be interpreted so that

$$D^r I(v) = \left[\frac{d^r}{dv^r} I(v)\right]_{v=\xi}. \tag{13}$$

Equation (11) then becomes

$$\Theta \exp\left[\left\{\frac{h(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}} - \xi\right\} D\right] I(v) = I(\xi). \tag{14}$$

This may now be solved by successive approximations.

The initial approximation is the large-sample normal approximation

$$h_0(w) = \xi\sqrt{(\Sigma\lambda_i w_i)}, \tag{15}$$

and we may write

$$h(w) = \xi\sqrt{(\Sigma\lambda_i w_i)} + h_1(w) + h_2(w) + \text{etc.}, \tag{16}$$

where $h_1(w)$ includes terms of order $1/f_i$, $h_2(w)$ terms of order $1/f_i^2$ and so on. For the moment we shall treat terms of the order $1/f_i^3$ as negligible. Then (14) gives

$$\Theta \exp\left[\left\{\frac{\xi\sqrt{(\Sigma\lambda_i w_i)}}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}} - \xi\right\} D\right] \exp\left[\left\{\frac{h_1(w) + h_2(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right\} D\right] I(v) = I(\xi), \tag{17}$$

i.e. $$\Theta \exp\left[\xi D\left\{\sqrt{\frac{\Sigma\lambda_i w_i}{\Sigma\lambda_i\sigma_i^2}} - 1\right\}\right]\left[1 + \frac{h_1(w)\,D}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}} + \left\{\frac{h_2(w)\,D}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}} + \frac{1}{2}\frac{h_1^2(w)\,D^2}{\Sigma\lambda_i\sigma_i^2}\right\} \ldots\right] I(v) = I(\xi). \tag{18}$$

Or, using (10),

$$\begin{aligned}&\left[\frac{h_1(\sigma^2)\,D}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}} + \Sigma\frac{\sigma_i^4\partial_i^2}{f_i}\exp\left(\xi D\left\{\sqrt{\frac{\Sigma\lambda_i w_i}{\Sigma\lambda_i\sigma_i^2}} - 1\right\}\right)\right] I(v)\\ &\quad + \left[\frac{h_2(\sigma^2)\,D}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}} + \frac{1}{2}\frac{h_1^2(\sigma^2)\,D^2}{\Sigma\lambda_i\sigma_i^2} + \Sigma\frac{\sigma_i^4\partial_i^2}{f_i}\exp\left(\xi D\left\{\sqrt{\frac{\Sigma\lambda_i w_i}{\Sigma\lambda_i\sigma_i^2}} - 1\right\}\right)\frac{h_1(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right.\\ &\quad + \left.\left\{\frac{4}{3}\Sigma\frac{\sigma_i^6\partial_i^3}{f_i^2} + \frac{1}{2}\left(\Sigma\frac{\sigma_i^4\partial_i^2}{f_i}\right)^2\right\}\exp\left(\xi D\left\{\sqrt{\frac{\Sigma\lambda_i w_i}{\Sigma\lambda_i\sigma_i^2}} - 1\right\}\right)\right] I(v) = 0.\end{aligned} \tag{19}$$

The equation of the first order term to zero gives

$$h_1(\sigma^2) = \frac{\xi(1+\xi^2)}{4}\frac{\left(\Sigma\frac{\lambda_i^2\sigma_i^4}{f_i}\right)}{(\Sigma\lambda_i\sigma_i^2)^{\frac{3}{2}}}. \tag{20}$$

* Equation (11) can also be expressed as an integral equation and this form may be necessary for providing numerical values where the f_i are very small.

This can then be substituted in the second-order term which, when equated to zero, will give $h_2(\sigma^2)$. The process may obviously be extended to higher orders, although the expressions become so complex that a slightly different procedure has then been found to be preferable. To terms of order $1/f_i^2$ our solution is

$$h(s^2) = \xi\sqrt{(\Sigma\lambda_i s_i^2)}\left[1+\frac{(1+\xi^2)}{4}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i}\right)}{(\Sigma\lambda_i s_i^2)^2}-\frac{(1+\xi^2)}{2}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i^2}\right)}{(\Sigma\lambda_i s_i^2)^2}\right.$$
$$\left.+\frac{(3+5\xi^2+\xi^4)}{3}\frac{\left(\Sigma\frac{\lambda_i^3 s_i^6}{f_i^2}\right)}{(\Sigma\lambda_i s_i^2)^3}-\frac{(15+32\xi^2+9\xi^4)}{32}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i}\right)^2}{(\Sigma\lambda_i s_i^2)^4}\right]. \tag{21}$$

It may be noted that in the particular case $k = 1$, this reduces, as it should, to the already known expansion of the deviate of the straightforward 'Student' distribution (Fisher, 1941, p. 151), viz.

$$t_P = \xi\left[1+\frac{(1+\xi^2)}{4f}+\frac{(3+16\xi^2+5\xi^4)}{96f^2}+\text{etc.}\right]. \tag{22}$$

It is proposed in another communication to give tables of $h(s^2)$ based on the expansion (21) carried to some further terms.

4. *Discussion of a non-series approximation.* It will be recalled that in Gosset's original approach to the single sample problem ('Student', 1908) his initial step was to note that the first four moments of the distribution of s^2 were consistent with the assumption that the distribution could be represented by a Pearson Type III curve. He was fortunate in this way to rediscover a distribution which had already been found by Helmert, as this permitted him to go on to the derivation of the t-distribution. In our present case, as in many others arising naturally in statistical work, we are led to consider, instead of s^2, a linear function $\Sigma\lambda_i s_i^2$ of several s_i^2. If this linear function were distributed in a Pearson Type III distribution a whole range of new problems could be dealt with by well-established theory. However, in general, we do not have this good fortune. For $\Sigma\lambda_i s_i^2$ is of the form $\Sigma a_i\chi_i^2$, where $a_i = \lambda_i\sigma_i^2/f_i$, and the distribution of this quantity is only of Type III if all the a_i, except one, are zero, or if all the a_i happen to be equal.

Nevertheless, for practical purposes an *approximation* to the distribution of $\Sigma\lambda_i s_i^2$, using a Type III curve with start, mean and variance suitably adjusted, can still be useful. In two previous papers (Welch, 1936, 1938) I have employed this method to obtain numerical comparisons of the merits of different statistical procedures, where full calculations with the true distributions would have been unduly laborious. The method of determining the constants in the approximation was given for the case $k = 2$ in the first of these papers and is as follows.

If $z = (a\chi_1^2 + b\chi_2^2)$, and the approximate distribution curve is written in the form

$$p(z)\,dz = \frac{1}{\Gamma(\frac{1}{2}f)}e^{-\frac{1}{2}(z/g)}\left(\frac{z}{2g}\right)^{\frac{1}{2}f-1}d\left(\frac{z}{2g}\right), \tag{23}$$

then making the first two moments of (23) agree with the true moments of z, we find

$$f = \frac{(af_1+bf_2)^2}{a^2f_1+b^2f_2},\quad g = \frac{a^2f_1+b^2f_2}{af_1+bf_2}. \tag{24}$$

Phrasing the matter rather differently, we can say that z/g is approximately distributed as

χ^2 with degrees of freedom f. Of course f, given by (24), will in general be fractional, but the letter used to designate this quantity was chosen, and the term 'effective degrees of freedom' has been used, because by doing so we can appeal immediately to a considerable body of further theoretical results.

In particular we can say that the criterion

$$v = \frac{(y-\eta)}{\sqrt{(\lambda_1 s_1^2 + \lambda_2 s_2^2)}} \tag{25}$$

follows approximately the 'Student' t-distribution with degrees of freedom

$$f = \frac{(\lambda_1\sigma_1^2 + \lambda_2\sigma_2^2)^2}{\frac{\lambda_1^2\sigma_1^4}{f_1} + \frac{\lambda_2^2\sigma_2^4}{f_2}}. \tag{26}$$

More generally, when k is not restricted to 2, the same line of argument leads us to say that the criterion

$$v = \frac{(y-\eta)}{\sqrt{(\Sigma\lambda_i s_i^2)}} \tag{27}$$

is approximately distributed as 'Student's' t with degrees of freedom

$$f = \frac{(\Sigma\lambda_i\sigma_i^2)^2}{\Sigma\frac{\lambda_i^2\sigma_i^4}{f_i}}. \tag{28}$$

Not knowing the σ_i's in (28), there are several ways in which we may now proceed, depending on what weight we may be willing to attach to any vague *a priori* notions we may possess of their *relative* magnitudes (cf. Welch, 1938). If we are not willing to assume anything, perhaps the best choice is

$$f = \frac{(\Sigma\lambda_i s_i^2)^2 - 2\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i+2}\right)}{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i+2}\right)}. \tag{29}$$

It may be shown that the numerator of (29) has, in repeated samples, an average value $(\Sigma\lambda_i\sigma_i^2)^2$, and the denominator has average value $\Sigma\lambda_i^2\sigma_i^4/f_i$. In a certain sense, therefore, (29) is a fair estimate of (28).

To sum up, then, the interpretation of y as an estimate of η, using the present type of approximation involves only the reference of the criterion (27) to tables of the 'Student' distribution, entered with degrees of freedom given by (29).

Some further light is now thrown on this procedure by the expansion for the exact solution of our problem derived in the preceding section. For the implications of referring v to the 'Student' distribution may be seen by substituting f from (29) into the expansion (22) of the 'Student' deviate. On doing this and then expanding in powers of $1/f_i$ it is found that, in effect, our approximation corresponds to assuming that

$$h(s^2) = \xi\sqrt{(\Sigma\lambda_i s_i^2)}\left[1 + \frac{(1+\xi^2)}{4}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i}\right)}{(\Sigma\lambda_i s_i^2)^2} - \frac{(1+\xi^2)}{2}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i^2}\right)}{(\Sigma\lambda_i s_i^2)^2} + \frac{(51+64\xi^2+5\xi^4)}{96}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i}\right)^2}{(\Sigma\lambda_i s_i^2)^4} + \dots\right], \tag{30}$$

whereas, in fact, the true solution is given by (21). Comparison shows that we have exact

agreement to terms of order $1/f_i$ and in the first of the quadratic terms. To the second order the difference between the expressions in square brackets in equations (21) and (30) is

$$\frac{(3+5\xi^2+\xi^4)}{3}\left\{\frac{\left(\Sigma\frac{\lambda_i^3 s_i^6}{f_i^2}\right)}{(\Sigma\lambda_i s_i^2)^3}-\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i}\right)^2}{(\Sigma\lambda_i s_i^2)^4}\right\}. \tag{31}$$

This difference vanishes if any one of the s_i^2 is overwhelmingly larger than all the others, or if s_i^2 is proportional to f_i/λ_i. It appears that, in general, the difference is not likely to be large. We have, therefore, found some justification for using the Type III approximation in the present case.

The above comparison has been made on the basis of the series developments, but it should be borne in mind that approximations based on positive frequency functions, such as those falling under the Pearson system, usually provide a higher degree of accuracy than might appear from any consideration of expansions. Furthermore, they are apt to give an insight into the nature of the situation which may sometimes be lost in working out the details of exact solutions. In the present case I feel that the comparison of this section serves to give added confidence in the exact solution,* which I have put forward in the previous two sections, quite as much as it demonstrates the value of the approximate method.

5. *Further discussion.* In comparing the present contribution with other work on the subject, the essential point to notice is the averaging process involved in equation (5). We are not trying here to make probability statements valid for *fixed* s_i^2, but are averaging over the joint probability distribution of the s_i^2, taking into account, therefore, the different values which can arise by chance in sampling from populations with fixed σ_i^2.

This averaging over the joint distribution of the s_i^2 is parallel to the step taken in Section III of Gosset's original memoir (1908) where, in effect, he starts with the distribution of t for samples with *fixed* s and then averages over the distribution of s which he has already derived earlier. He thus arrives at the unrestricted distribution of t (or, more strictly, of a quantity z, which is equal to t multiplied by a constant). This distribution forms the basis of the significance tests which he illustrates in his Section IX and of the method of deriving interval estimates for the population mean which he outlines in his Section VIII.

In the present paper the parallelism with Gosset's work may be obscured to some extent by the fact that we do not from the outset seek the probability distribution of some pivotal quantity like t, explicitly expressed. It so happens that we are able to proceed to a method of deriving an expansion for the required probability level without making explicit reference to such a quantity. Nevertheless there remains the important resemblance with Gosset's development, in that we do not confine ourselves to samples with fixed s_i^2.

This procedure stands in sharp contrast to the formulation of the problem of comparing two means, favoured by R. A. Fisher (e.g. 1941) and H. Jeffreys (1940). These writers prefer a solution which they ascribe initially to W. U. Behrens (1929). Looked at from one point of view, Behrens's paper appears to contain some gross algebraical errors. Fisher and Jeffreys, however, develop lines of argument by means of which they claim that Behrens's solution is quite justified. It seems to me difficult to say how far (if at all) any of these arguments may have been in Behrens's mind when he wrote his paper and I shall not attempt to elucidate this question here. We may, however, permit ourselves one observation about the developments according to Fisher and Jeffreys.

* Exact in the sense that it is independent of the irrelevant population parameters σ_i^2.

Both these writers, at some stage, limit the field of their probability inferences to a subset in which the s_i^2 are regarded as *fixed*. In order to solve the problem on these lines Jeffreys introduces an *a priori* distribution function for the unknown σ_i, following his general philosophy for dealing with such questions. Fisher, on the other hand, arrives at the same answer by a special utilization of what he terms the *fiducial* distribution of σ_i.

Jeffreys's approach here does not raise any new issues to those who are familiar with the general body of his researches on statistical inference. Fisher's justification of Behrens's solution is perhaps of more immediate interest as it raises controversial points which are important more specifically in relation to our present topic of discussion. For although Fisher's approach has been very much criticized by a number of writers, starting with M. S. Bartlett (1936), the critics have not wished to throw doubt on the whole body of results which Fisher includes under the heading of fiducial inference. The criticism has been for the most part selective, directed mainly at the way in which so-called *simultaneous* fiducial distributions of several parameters have been defined and manipulated.

I have, myself, quite definite views on these questions (particularly on the usage of the word 'fiducial') but do not feel that I need express them at any great length here. I disagree with Fisher, but this divergence of opinion must already have become apparent in the way I have defined the field within which I make my probability inferences about η. It appears to me to be quite artificial to restrict our view to one which, even in a limited sense, fixes s_i^2. It is true that, in the two-sample problem, we have to draw our inferences from the unique pair of samples observed, or, more precisely, from the statistics $\bar{x}_1$, $\bar{x}_2$, s_1^2 and s_2^2 which they provide. These statistics are our only *data* for the purpose of making inferences, but we add something to these data in the *interpretation* when we regard the samples as being drawn randomly from hypothetical normal populations. Once having embarked on this method of interpretation, we should stick to it consistently throughout. The sampling variations of s_i^2 should be taken into account only by a direct use of the probability distributions as given by our equation (1) and not by any inversion such as is involved in Fisher's conception of the fiducial distribution of σ_i^2. As we have seen, it is quite possible to make probability statements about the difference between the population means without making any reference whatever either to inverse probability or to fiducial distributions.

The distinction between the procedure which Fisher advocates and one which averages over the s^2 distributions has, of course, been stressed by most of the writers who have contributed papers on the subject, from whatever viewpoint (e.g. Bartlett, 1936, p. 566, and Yates, 1939.) What has been lacking hitherto, however, is a solution, analogous to Gosset's single sample solution, which makes complete use of the information contained in the data provided. Bartlett indicated one particular way in which probability inferences about the difference between two population means might be made, but was careful to point out that the problem of making the best possible inferences (in the theoretical sense of utilizing all the information in the data to its full extent) was still an open one. There has indeed been some doubt expressed whether a fully satisfactory solution from this point of view existed at all. I believe, however, that the one I advance above in equation (11), and develop in equation (21), meets all the requirements that one can reasonably expect.

Whatever conclusion the reader may come to on these matters, however, he will probably wish to know how, in the numerical details, this solution will differ from that of Behrens. This will be more easily seen when some tables become available, but fortunately certain

comparisons can already be made. For Fisher (1941, p. 155) has provided a series expansion of the Behrens solution. In our notation, and with $k=2$, this may be written, to order $1/f_i$, as follows:

$$h(s^2) = \xi\sqrt{(\lambda_1 s_1^2+\lambda_2 s_2^2)}\left[1+\frac{(1+\xi^2)}{4}\frac{\left(\frac{\lambda_1^2 s_1^4}{f_1}+\frac{\lambda_2^2 s_2^4}{f_2}\right)}{(\lambda_1 s_1^2+\lambda_2 s_2^2)^2}+\left(\frac{1}{f_1}+\frac{1}{f_2}\right)\frac{\lambda_1\lambda_2 s_1^2 s_2^2}{(\lambda_1 s_1^2+\lambda_2 s_2^2)^2}\right]. \tag{32}$$

Even to this order, this differs from our equation (21) in the inclusion of an extra term. In other words, although the two solutions are the same when samples are large enough to adopt the large-sample normal approximation, they differ immediately we take into account the first corrective term, i.e. they differ as soon as we begin to attach any importance to 'Studentization'.

6. *An interval estimate for η.* We have shown in §§ 2 and 3 how to calculate a value $h(s_1^2, s_2^2, \ldots, s_k^2, P)$, depending on the observed variances $s_1^2, s_2^2, \ldots, s_k^2$, such that the probability is P that $(y-\eta) < h(s_1^2, s_2^2, \ldots, s_k^2, P)$. This provides a method of testing the consistency of an observed y with a prescribed value η.

When the question is not whether any particular given η is contradicted by the data, but rather one of estimating η and at the same time of providing a measure of the uncertainty of the estimate, the further step required is immediate. For, as in the case of a single sample, the order of the words in our probability statement can be changed so that it becomes—the probability is P that η is greater than $\{y-h(s_1^2, s_2^2, s_k^2, P)\}$. An interval estimate for η is then obtained by taking two levels P_1 and P_2 for P. Thus the probability is (P_1-P_2) that η lies between $\{y-h(s_1^2, s_2^2, \ldots, s_k^2, P_1)\}$ and $\{y-h(s_1^2, s_2^2, \ldots, s_k^2, P_2)\}$.

If $P_2 = (1-P_1)$ the range will be symmetrically placed about y. Thus, for example, if $P_1 = 0{\cdot}95$ and $P_2 = 0{\cdot}05$, the chance will be 90 % that η lies within the range

$$\eta \pm 1{\cdot}6449\sqrt{(\Sigma\lambda_i s_i^2)}\left[1+\frac{1+(1{\cdot}6449)^2}{4}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i}\right)}{(\Sigma\lambda_i s_i^2)^2}+\text{etc.}\right]. \tag{33}$$

It may be noted, incidentally, that this range is always narrower than similar ranges calculated from Behrens's solution.

REFERENCES

BARTLETT, M. S. (1936). The information available in small samples. *Proc. Camb. Phil. Soc.* **32**, 560–6.

BEHRENS, W. U. (1929). Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtungen. *Landw. Jb.* **68**, 807–37.

FISHER, R. A. (1941). The asymptotic approach to Behrens's integral, with further tables for the *d* test of significance. *Ann. Eugen., Lond.*, **11**, 141–72.

JEFFREYS, H. (1940). Note on the Behrens-Fisher formula. *Ann. Eugen., Lond.*, **10**, 48–51.

'STUDENT' (1908). The probable error of a mean. *Biometrika*, **6**, 1–25.

WELCH, B. L. (1936). Specification of rules for rejecting too variable a product, with particular reference to an electric lamp problem. *J. Roy. Statist. Soc. Suppl.* **3**, 29–48.

WELCH, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, **29**, 350–62.

YATES, F. (1939). An apparent inconsistency arising from tests of significance based on fiducial distributions of unknown parameters. *Proc. Camb. Phil. Soc.* **35**, 579–91.

THE DISTRIBUTION OF KENDALL'S τ COEFFICIENT OF RANK CORRELATION IN RANKINGS CONTAINING TIES

By G. P. SILLITTO, Ph.D., B.Sc., *Research Department, I.C.I. (Explosives Ltd.)*

A new coefficient of rank correlation has been described by Kendall (1938, 1942, 1943) and denoted by him as τ. This coefficient has advantages over Spearman's ρ in respect of the smoothness of its distribution and the rapidity with which it approaches normality, thus facilitating significance testing, and in being readily adapted to cases of partial rank correlation.

The distribution of τ has been worked out by Kendall (1938, 1943) for cases in which neither ranking contains members which are graded equal, i.e. rankings containing no 'ties'. It is the purpose of the present paper to deal with cases, which frequently arise in practice, in which ties occur in one of the two rankings. The method is a generalization of that of Kendall and will be given in some detail for the case of tied pairs, while the results of further generalization to multiplet ties will be indicated without detailed proof, which can in all cases be effected simply on the lines indicated.

Definition of τ for rankings containing ties

In counting the 'score' of a pair of rankings, by the methods suggested by Kendall, each member is compared with the other members of the same ranking, and additions to or subtractions from the score are made depending on whether it is smaller or greater in each case. If some members are ranked equal then it is proposed that no change be made in the score in comparing them. This obviously accords with the intuitive aspects of ranking. Thus in the pair of rankings following, the score is +8:

$$\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 3 & 5 & 6 & 3 \end{array}$$

The maximum score possible is thus obviously reduced by the presence of ties, and it is evident that the presence of each tied pair reduces the maximum possible score by unity, so that it becomes $\frac{1}{2}n(n-1)-p_2$ for the case of a ranking of n members containing p_2 pairs. Thus for such a ranking τ would be defined as

$$\tau = 2S/\{n(n-1)-2p_2\},$$

where S is the observed score.

Generally, each r-tuplet tie reduces the maximum possible score by $\frac{1}{2}r(r-1)$* so that for a ranking of n members containing p_2 pairs, p_3 triplets, ..., p_r r-tuplets,

$$\tau = \frac{2S}{n(n-1)-2p_2-6p_3-\ldots-r(r-1)p_r}.$$

The sum of the frequencies of the possible scores

When no ties are present, each permutation of the n members produces a possible score so that there are in all $n!$ possible scores. When ties are present they decrease the number of possible permutations of an assigned set of members, but, on the other hand, they give rise

* This result has been given by Kendall (1945).

to further families of scores due to the different places in the ranking which can be occupied by the tied members. Thus, for instance, the rankings

113456, 122456, 123356, 123446, 123455

all give rise to the maximum score, 14.

Considering any assigned ranking, the number of possible permutations with p_2 pairs present is $n!/2^{p_2}$, or if there are in addition p_3 triplets, ..., p_r r-tuplets, $n!/(2!)^{p_2}(3!)^{p_3}\dots(r!)^{p_r}$.

The distribution of scores of an assigned set of ranks will be referred to as the basic distribution for the type of ranking concerned, since consideration of the possible ways of assigning the p_2 pairs, p_3 triplets, etc., among the members of the ranking has only the effect of multiplying the frequency of each score by a constant factor. This factor is the number of ways of distributing the $p_1+p_2+p_3+\dots+p_r$ ranks among the n members. This is the number of possible permutations

$$\frac{(p_1+p_2+p_3+\dots+p_r)!}{p_1!p_2!p_3!\dots p_r!}.$$

BASIC FREQUENCY DISTRIBUTIONS OF THE SCORES

The basic frequency distributions can be established by an extension of the methods given by Kendall. Considering first the case of tied pairs the frequency function of the basic distribution of the scores may be written $f(S, n, p_2)$, where p_2 is the number of pairs. The frequency generating function is then $\sum_j f(S_j, n, p_2)\,t^{S_j}$. Now consider the addition of another tied pair, with a greater ranking than any of the existing ranks. If it is added to the extreme left of the ranking it adds $-2n$ to the score. Moving one of the pair one place to the right adds 2 to this new score; bringing the other added member up to it adds another 2. Starting again with both the new members on the extreme left, movement of one of them two places to the right adds 4 to the new score, bringing the other up to it in two steps each of one place, adds successively a further 2 and 4. Proceeding in this way all possible additions to the old score which may be brought about by the addition of a tied pair of new members are represented by the array

$$\begin{array}{cccccc}
-2n & -(2n-2) & -(2n-4) & -(2n-6) & \dots & 0 \\
 & -(2n-4) & -(2n-6) & -(2n-8) & & +2 \\
 & & -(2n-8) & -(2n-10) & & +4 \\
 & & & -(2n-12) & & \dots \\
 & & & & & \dots \\
 & & & & & +2n
\end{array}$$

Thus the addition of a new tied pair has the effect of multiplying the frequency generating function by

$$\{t^{-2n}+(t^{-(2n-2)}+t^{-(2n-4)})+(t^{-(2n-4)}+t^{-(2n-6)}+t^{-(2n-8)})+\dots+(t^0+t^2+\dots+t^{2n})\}.$$

The addition of a single new member to the ranking has the effect of multiplying the frequency generating function by

$$\{t^{-n}+t^{-(n-2)}+t^{-(n-4)}+\dots+t^{n}\}$$

as shown by Kendall, the presence of tied pairs in the existing ranking having no effect.

With these two recurrence relations there is no difficulty in drawing up a table of basic frequency distributions for tied pairs as exemplified in Table 1, in which only positive values of the score are shown, negative values being obtainable by symmetry.

Table 1. *Distribution of the score S for values of n from 3 to 7, and for rankings containing p_2 pairs of members ranked equal (only positive half of symmetrical distribution)*

n	p_2	Values of S																					
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
3	.	.	2	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3	1	1	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4	.	6	.	5	.	3	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4	1	.	3	.	2	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4	2	2	.	1	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
5	.	22	.	20	.	15	.	9	.	4	.	1	.	.	.	.	.	.	.	.	.	.	.
5	1	.	11	.	9	.	6	.	3	.	1	.	.	.	.	.	.	.	.	.	.	.	.
5	2	6	.	5	.	4	.	2	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.
6	.	.	101	.	90	.	71	.	49	.	29	.	14	.	5	.	1	.	.	.	.	.	.
6	1	52	.	49	.	41	.	30	.	19	.	10	.	4	.	1	.	.	.	.	.	.	.
6	2	.	26	.	23	.	18	.	12	.	7	.	3	.	1	.	.	.	.	.	.	.	.
6	3	14	.	12	.	11	.	7	.	5	.	2	.	1	.	.	.	.	.	.	.	.	.
7	.	.	573	.	531	.	455	.	359	.	259	.	169	.	98	.	49	.	20	.	6	.	1
7	1	202	.	281	.	250	.	205	.	154	.	105	.	64	.	34	.	15	.	5	.	1	.
7	2	.	146	.	135	.	115	.	90	.	64	.	41	.	23	.	11	.	4	.	1	.	.
7	3	74	.	72	.	63	.	52	.	38	.	26	.	15	.	8	.	3	.	1	.	.	.

Before the construction of the table has proceeded far, however, it becomes evident that there is a recurrence relation between individual frequencies for any given value of n, such that the frequency of any score S_j for p_2 pairs is the sum of the frequencies of $S_j - 1$ and $S_j + 1$ for $p_2 + 1$ pairs. This obviously arises from the fact that if two members ranked equal, say rth, in a ranking with $p_2 + 1$ pairs are subsequently distinguished and given rankings r and $r+1$, this will increase the score by unity if the $(r+1)$ member falls after the (r) member when the ranking is arrayed against another ranking in the natural order $1, 2, 3, \ldots, n$, and reduce it by unity if the other member of the pair becomes the $(r+1)$th; and these two possibilities complete the ways of forming a ranking with p_2 pairs from one with $p_2 + 1$ pairs.

This simple relationship, which may be written

$$f(S_j, n, p_2) = f(S_j + 1, n, p_2 + 1) + f(S_j - 1, n, p_2 + 1),$$

or taking another way of writing the basic distribution function

$$\phi(S_j, p_1, p_2) = \phi(S_j + 1, p_1 - 2, p_2 + 1) + \phi(S_j - 1, p_1 - 2, p_2 + 1), \tag{1}$$

p_1 being the number of members not in tied pairs, is of great assistance in tabulating the frequency distribution, and will be used below to establish the formula for the variance of S. It can be generalized to cover the effect of increasing the number of r-tuplets, when it becomes

$$\begin{aligned}\phi(S_j, p_1, p_2, \ldots, p_{r-1}, p_r) = \; & \phi(S_j - r - 1, p_1 - 1, p_2, \ldots, p_{r-1} - 1, p_r + 1) \\ & + \phi(S_j - r - 3, p_1 - 1, p_2, \ldots, p_{r-1} - 1, p_r + 1) \\ & + \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ & + \phi(S_j + r - 1, p_1 - 1, p_2, \ldots, p_{r-1} - 1, p_r + 1).\end{aligned} \tag{2}$$

FREQUENCY AND PROBABILITY DISTRIBUTIONS OF THE SCORE S

From a table of basic frequency distributions such as Table 1 the construction of a table showing the probability of attaining or exceeding an observed value of S by chance from an uncorrelated pair of rankings can obviously be constructed, and Table 2 shows such probabilities (positive S only, negative values obtainable by symmetry) for values of n up to 10, and all possible numbers p_2 of paired and p_3 of triplet ties.

THE VARIANCE OF S

The variance of S when ties are present can be readily derived by using the recurrence relations given above and the value given by Kendall for the case of no ties. For the case of tied pairs consider

$$\begin{aligned}(S+1)^2\phi(S+1, p_1-2, p_2+1)+(S-1)^2\phi(S-1, p_1-2, p_2+1)\\ = S^2\{\phi(S+1, p_1-2, p_2+1)+\phi(S-1, p_1-2, p_2+1)\}\\ +2\{(S+1)\phi(S+1, p_1-2, p_2+1)-(S-1)\phi(S-1, p_1-2, p_2+1)\}\\ -\{\phi(S+1, p_1-2, p_2+1)+\phi(S-1, p_1-2, p_2+1)\}.\end{aligned}$$

If now both sides of this equation are summed over all values of S, the terms on the left-hand side become

$$\frac{n!}{(2!)^{p_2+1}}\operatorname{var}\phi(S, p_1-2, p_2+1).$$

The first terms on the right-hand side become by virtue of the recurrence relation (1)

$$\frac{n!}{(2!)^{p_2}}\operatorname{var}\phi(S, p_1, p_2),$$

the second vanishes through the symmetry of the distribution, while the third becomes

$$-\frac{2.n!}{(2!)^{p_2+1}}.$$

Hence there is obtained

$$\operatorname{var}\phi(S, p_1-2, p_2+1) = \operatorname{var}\phi(S, p_1, p_2)-1,$$

and so

$$\begin{aligned}\operatorname{var}\phi(S, n-2p_2, p_2) &= \operatorname{var}\phi(S, n)-p_2\\ &= \frac{n(n-1)(2n+5)}{18}-p_2,\end{aligned}$$

using Kendall's result.

These results can also be generalized, using equation (2), to deal with multiplet ties, obtaining

$$\operatorname{var}\phi(S, p_1-1, \ldots, p_{r-1}-1, p_r+1) = \operatorname{var}\phi(S, p_1, \ldots, p_{r-1}, p_r)-(r^2-1)/3,$$

and
$$\operatorname{var}\phi(S, p_1, p_2, \ldots, p_r) = \frac{n(n-1)(2n+5)}{18}-p_2-\frac{3+8}{3}p_3-\ldots-\frac{3+8+\ldots+(r^2-1)}{3}p_r.$$

It is obvious from these equations for multiplet ties that for any given number of ties of each multiplicity the variance will tend towards that of the system without any ties as n increases.

APPLICATION TO A PRACTICAL CASE

The following results were obtained in a practical case in which two different tests were carried out on one each of a set of products. The problem is to determine the degree of relationship between the results of the two tests. It is also an instance of an occurrence

which arises at times in practice, in which some of the results are 'off the scale' of measurement with respect to one of the tests; these, twelve in all, have been given a tied ranking of 18.

Test A	40·80	41·70	36·75	37·55	29·40	25·20	26·75	28·45	26·85	26·35
Test B	1·5	1·5	1·5	2·5	3·5	10	2·5	2·5	2·5	6·5
Test A	21·40	19·65	18·95	22·90	22·80	20·25	24·45	22·70	26·50	—
Test B	>10	>10	>10	>10	>10	>10	>10	>10	>10	—
Test A	22·00	27·50	23·75	30·80	21·00	27·10	22·10	19·25	25·45	24·10
Test B	>10	3·5	1·5	2·5	7	6·5	7·5	9	3·5	>10

Ranking the results according to their order in test A (from highest values to lowest) there is obtained

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	5	1	5	10	5	10	13	5	5	18	13	10	18
16	17	18	19	20	21	22	23	24	25	26	27	28	29	
18	18	1	18	18	18	16	18	18	15	18	18	17	18	

The lower ranking has 1 pair, 1 triplet, 1 quadruplet, 1 quintuplet and one 12-member multiplet. The maximum possible score is $\frac{29 \times 28}{2} - 1 - 3 - 6 - 10 - 66 = 320$ in such a ranking. The actual score is $+212$, giving $r = \frac{212}{320} = 0 \cdot 6625$. The variance of the distribution of the scores obtained with such rankings in the case of no correlation is

$$\frac{29.28.63}{18} - 1 - \frac{11}{3} - \frac{26}{3} - \frac{50}{3} - \frac{638}{3} = 2599 \cdot 33.$$

Hence the probability of obtaining a score of 212 or more from an uncorrelated pair of such rankings corresponds to the probability of a normal variate attaining or exceeding $\frac{212}{\sqrt{(2599 \cdot 33)}} = 4 \cdot 158$ times its standard deviation.

REFERENCES

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**, 81.

Kendall, M. G. (1942). Partial rank correlation. *Biometrika*, **32**, 277.

Kendall, M. G. (1943). *The Advanced Theory of Statistics*, **1**, chapter 16, Rank Correlation. London: C. Griffin and Co. Ltd.

Kendall, M. G. (1945). The treatment of ties in ranking problems. *Biometrika*, **33**, 239.

THE USE OF RANGE IN PLACE OF STANDARD DEVIATION IN THE t-TEST

By E. LORD, B.Sc., *Shirley Institute, Didsbury, Manchester*

CONTENTS

(i) Introduction

The difference between highest and lowest values has always been recognized as a general indication of the variability of quantitative data. It was not, however, until 1925 that attention began to be focused upon the range as a useful statistical tool. In his paper 'On the extreme individuals and the range in samples taken from a normal population', Tippett (1925) obtained an expression for the mean value of the range in repeated random samples, and computed its value in terms of the population standard deviation for samples of size $n = 2$ to $n = 1000$. He also gave numerical approximations to the values of the moments of the range for fairly large samples.

The work was taken up by E. S. Pearson (1926), who determined numerically the exact values of the moments of the random sampling distribution of the range for small samples of size $n \leqslant 6$, and also approximations to their values for samples of medium size. In a subsequent paper E. S. Pearson (1932) tabulated the upper and lower percentage limits for the distribution of the range from frequency curves fitted with the values of the moment coefficients taken from both of the earlier papers cited above.

The next advance was the determination of a general expression for the distribution of the range in samples of n random values from any population by McKay & Pearson (1933). For the normal population, only in the case of $n = 3$ was it found possible to obtain a fairly simple analytical form. (The distribution for $n = 2$ is, of course, well known, taking the form of the positive half of a normal curve.)

Hartley (1942) later determined an expression for the probability integral of the range and, with Pearson (1942), tabulated this for the normal population for samples between $n = 2$ and $n = 20$. This latter paper also contains a table of several percentage limits of the range in samples from a normal population. These limits are derived from the numerical values of the probability integrals and replace the approximate values previously given by Pearson referred to above.

Tippett (1925) and Pearson (1932) have pointed out that although the total range in a sample may be used for the purpose of estimating the population value of the standard

deviation, a more efficient measure may be obtained by dividing the sample into random subgroups of equal size and using the mean range of the subgroups in place of the total range. The efficiency of range estimates of standard deviation is, of course, always less than that of root-mean-square estimates, but the work of Davies & Pearson (1934) and Pearson & Haines (1935) indicates that information is not discarded to any serious extent providing that the number of observations in the subsamples is not greatly in excess of about 10.

As a result of the work outlined above, the range is now of considerable importance in many fields, especially in industrial quality control, where its simplicity has enabled it to be extensively and easily applied to the measurement of fluctuations in the variability of quality of a manufactured article or material.

In the present paper an investigation is made of the use of range estimates of standard deviation in the consideration of the statistical significance of deviations of sample means in normal random sampling theory. This use of range estimates of standard deviation is analogous to the use of root-mean-square estimates in the well-known t-test. Tables are given, at several probability levels, and these may be employed in determining the statistical significance of either the deviation of a sample mean from some fixed or hypothetical population value, or the difference between the means of two samples. These tables may also be used for obtaining rapid estimates of the accuracy of a sample mean from the variation within the sample as measured by the range. The use of range, in place of root-mean-square estimates of standard deviation, in this modified form of the t-test necessarily entails some loss of precision. It will, however, be shown in a future paper that this reduction in accuracy is negligible for all practical purposes. Furthermore, this slight disadvantage of the new test is compensated by its greater simplicity, involving a reduced amount of computing compared with the usual t-test.

The range test is suitable for application to many problems frequently encountered in the treatment of various types of experimental data and in considering the mean character value in small samples in biological experiments. In the industrial field, the range test may be used for detecting changes in mean quality level, especially where the variation is not under strict statistical control or is subject to secular changes, or for determining whether the average level of a batch determined from a sample is in accordance with specification demands. A number of these problems are covered in the examples given at the end of the paper.

(ii) The t-test*

In testing the significance of the deviation of a sample mean $\bar{x}$ from an assumed population value ξ, use is made of the ratio

$$t = \frac{|\bar{x}-\xi|}{s/\sqrt{N}}, \tag{1}$$

where N is the size of the sample and s is the root-mean-square estimate of the population standard deviation determined from the sample. In applying this ratio it is assumed that the N values form a random sample from a normal population of which the mean is ξ, standard deviation σ and the distribution of values of x is given by

$$p(x) = \frac{1}{\sqrt{(2\pi)}\,\sigma} e^{-\frac{(x-\xi)^2}{2\sigma^2}}. \tag{2}$$

* 'Student' (1908), R. A. Fisher (1925).

More generally t may be defined as the ratio

$$t = x/s, \tag{3}$$

where x and s are statistically independent, x being a quantity distributed normally about a mean of zero and s a root-mean-square estimate based on ν degrees of freedom of the standard error of x. Although the use of the tables of the probability integral of t enables the most efficient tests to be made of the various forms of the so-called 'Student's Hypothesis', occasions frequently arise when more rapid tests are desirable, especially if accompanied by only inappreciable loss of accuracy. The calculation of s, depending upon the squaring of numerical quantities, entails a certain amount of labour, especially if tables of squares or a calculating machine are not available. The use of the range, or the mean range determined from random subgroups in a sample, enables very rapid estimates to be made of the population value of the standard deviation σ. In the following section these range estimates are used in place of root-mean-square estimates in a modified form of the t-test.

(iii) THE MODIFIED TEST (u-TEST) BASED ON RANGE

Here we replace the s of 'Student's' ratio by an estimate of σ based on the range. Thus

$$u = u(m, n) = \frac{x}{\bar{w}(m, n)/d_n}, \tag{4}$$

where x is a quantity distributed normally about a mean of zero and $\bar{w}(m, n)$ is the mean value of m ranges w, obtained from m independent samples or subgroups, each containing n observations. The constant d_n, in a commonly used notation,* is the expected value of the range in samples of n, randomly selected from a normal population of unit standard deviation. The ratio $\bar{w}(m, n)/d_n$ is therefore an estimate of the standard error of x obtained from range and, as such, replaces the root-mean-square estimate s used in the ratio $t = x/s$.

Except for a few special cases, it has not been found possible to determine the analytical form of the distribution of u, but several tables of percentage points have been computed for use in testing the various statistical hypotheses normally covered by the t-test. The computation of these tables is considerably simplified by first determining the percentage points of the distribution of the subsidiary quantity

$$q = q(m, n) = \frac{u(m, n)}{d_n} = \frac{x}{\bar{w}(m, n)}, \tag{5}$$

and the multiplying by the corresponding value of d_n to obtain the percentage points of the u distribution.

To simplify the algebraic expressions in what follows, u, w and q will be written for $u(m, n)$, $\bar{w}(m, n)$ and $q(m, n)$ where no confusion is involved.

The distribution of both u and q are clearly independent of σ. Hence, without any loss of generality, σ may be taken equal to unity in considering the distributions. The distribution of x will therefore be defined by

$$p(x) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}. \tag{6}$$

* See, for example, Pearson (1935), pp. 84 and 90.

Furthermore, let the distribution of the range w in a sample of n be $y = p(w)$, that of $\bar{w}$ be $y = p(\bar{w})$, and that of q be $y = p(q)$. Then since x and $\bar{w}$ are defined to be statistically independent, we have the distribution of q given by

$$p(q) = \int p(\bar{w})\, p(x)\, d\bar{w}\, dx, \tag{7}$$

where the integral is to be evaluated over the field of all values of x and $\bar{w}$ subject to the relation (5) and to the conditions:

$$-\infty < x < \infty, \quad 0 \leqslant \bar{w} < \infty. \tag{8}$$

Since x is distributed symmetrically about zero, and $\bar{w}$ is independent of x, the ratio q is also symmetrically distributed about zero. Let q_α be the value of q such that α is the chance that $|q| \geqslant q_\alpha$. The quantity α represents the total area of the two equal tails of the distribution lying outside deviations $\pm q_\alpha$, and we have

$$(1-\alpha) = 2\int_0^{q_\alpha} p(q)\, dq. \tag{9}$$

Alternatively, from (6), (7) and (8), this may be written in the form

$$\begin{aligned} 1-\alpha &= \int_0^\infty \left\{ p(\bar{w}) \int_{-\bar{w}q_\alpha}^{+\bar{w}q_\alpha} p(x)\, dx \right\} d\bar{w} \\ &= 2\int_0^\infty \left\{ p(\bar{w}) \int_0^{\bar{w}q_\alpha} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}\, dx \right\} d\bar{w}. \end{aligned} \tag{10}$$

Except for a few cases in which the analytical form of the distribution of u has been obtained, equation (10) has been used to compute values of q_α and, hence, the percentage points of u for values of $\alpha = 0{\cdot}10$, $0{\cdot}05$, $0{\cdot}02$, $0{\cdot}01$, $0{\cdot}002$ and $0{\cdot}001$, with values of n from 2 to 20 and values of m from 1 to 10, 15, 20, 30, 60 and 120.

The percentage points of $u = u(m, n)$ are first considered for the case when the estimate of standard deviation is based on the value of a single range of n random values (i.e. for $m = 1$). This treatment is followed by the case of $m = 2$, and finally consideration is given to the general case using estimates of standard deviation determined from the mean of m ranges each from an independent subgroup of n random values.

(iv) Computation of percentage points of the distribution of $u = u(1, n)$, i.e. case with $m = 1$

Throughout this section the estimates of standard deviation are all based upon the value of the range in a single set of n random values of the variate (thus $\bar{w} = w$). In the case of $n = 2$ and $n = 3$ analytical solutions are derived for the distributions from which the percentage points of q and u are calculated. For $n \geqslant 4$, percentage points in the neighbourhood of those desired are determined by quadrature methods, and the required points obtained from these by interpolation.

Special case $n = 2$, $m = 1$

The distribution of ranges (w) in samples of two random values from a normal population with unit standard deviation is the distribution of absolute differences between random pairs of variate values, and this may be easily shown to be

$$p(w)\, dw = \frac{1}{\sqrt{\pi}} e^{-\frac{1}{4}w^2}\, dw \quad (0 \leqslant w < \infty). \tag{11}$$

Since x and w are independent, it follows from (6) and (11) that their joint probability distribution is

$$p(w,x)\,dw\,dx = \frac{1}{\pi\sqrt{2}}e^{-\frac{1}{2}(x^2+\frac{1}{2}w^2)}\,dw\,dx. \tag{12}$$

Transforming to new variables, $q = x/w$ and w and noting that the Jacobian of the transformation

$$\frac{\partial(x,w)}{\partial(q,w)} = w,$$

the joint distribution of q and w is given by

$$p(q,w)\,dq\,dw = \frac{1}{\pi\sqrt{2}}e^{-\frac{1}{2}w^2(q^2+\frac{1}{2})}\,w\,dq\,dw. \tag{13}$$

To obtain the distribution of q it is necessary to integrate (13) over the whole field of w, from 0 to ∞. This gives

$$p(q)\,dq = \frac{dq}{\pi\sqrt{2}\,(q^2+\frac{1}{2})}. \tag{14}$$

Hence from (9) and (14) above, the percentage points of q are given by

$$\begin{aligned}(1-\alpha) &= \frac{2}{\pi\sqrt{2}}\int_0^{q_\alpha}\frac{dq}{(q^2+\frac{1}{2})}\\ &= \frac{2}{\pi}\tan^{-1}(\sqrt{2}\,q_\alpha),\end{aligned}$$

and hence

$$q_\alpha(1,2) = \frac{1}{\sqrt{2}}\tan\left\{\frac{\pi}{2}(1-\alpha)\right\} = \frac{1}{\sqrt{2}}\cot\left(\frac{\pi\alpha}{2}\right). \tag{15}$$

The values of q_α determined from (15), for the six values of α under consideration, are multiplied by $d_2 = 2/\sqrt{\pi}$ to give the required percentage points of the distribution of $u = u(1,2)$.

Special case $n = 3$, $m = 1$

For random samples of size $n = 3$ from a normal distribution with unit standard deviation, the distribution of the range has been found by McKay & Pearson (1933) and takes the form

$$p(w) = \frac{6}{\pi\sqrt{2}}e^{-\frac{1}{4}w^2}\int_0^{w/\sqrt{6}}e^{-\frac{1}{2}z^2}dz.$$

Again, since w and x are independent, their joint distribution is given by

$$p(w,x)\,dw\,dx = \frac{3}{\pi\sqrt{\pi}}e^{-\frac{1}{2}(x^2+\frac{1}{2}w^2)}\,dw\,dx\int_0^{w/\sqrt{6}}e^{-\frac{1}{2}z^2}\,dz. \tag{16}$$

Transforming to new variables $q = x/w$ and w, it follows from (16), since the Jacobian of the transformation is equal to w, that the joint distribution of q and w is given by

$$p(q,w)\,dq\,dw = \frac{3}{\pi\sqrt{\pi}}e^{-\frac{1}{2}w^2(q^2+\frac{1}{2})}\,w\,dw\,dq\int_0^{w/\sqrt{6}}e^{-\frac{1}{2}z^2}\,dz. \tag{17}$$

To obtain the distribution of q the expression in (17) has to be integrated over the whole field of w, from 0 to ∞. Thus

$$p(q) = \frac{3}{\pi\sqrt{\pi}} \int_0^\infty e^{-\frac{1}{2}w^2(q^2+\frac{1}{2})}\, w\,dw \int_0^{w/\sqrt{6}} e^{-\frac{1}{2}z^2}\,dz.$$

Putting $t = z\sqrt{6}$, the above may be written

$$\begin{aligned} p(q) &= \frac{3}{\pi\sqrt{(6\pi)}} \int_0^\infty e^{-\frac{1}{2}w^2(q^2+\frac{1}{2})}\, d(\tfrac{1}{2}w^2) \int_0^w e^{-t^2/12}\,dt \\ &= \frac{-3}{\pi\sqrt{(6\pi)}} \left[\frac{1}{(q^2+\frac{1}{2})} e^{-\frac{1}{2}w^2(q^2+\frac{1}{2})} \int_0^w e^{-t^2/12}\,dt \right]_{w=0}^{w=\infty} \\ &\quad + \frac{3}{\pi\sqrt{(6\pi)}\,(q^2+\frac{1}{2})} \int_0^\infty e^{-w^2/12} e^{-\frac{1}{2}w^2(q^2+\frac{1}{2})}\,dw. \end{aligned} \tag{18}$$

The first expression in (18) is clearly zero and hence

$$\begin{aligned} p(q) &= \frac{3}{\pi\sqrt{(6\pi)}\,(q^2+\frac{1}{2})} \int_0^\infty e^{-\frac{1}{2}w^2(q^2+\frac{2}{3})}\,dw \\ &= \frac{\sqrt{3}}{2\pi(q^2+\frac{1}{2})\,(q^2+\frac{2}{3})^{\frac{1}{2}}}. \end{aligned} \tag{19}$$

As before $$(1-\alpha) = \frac{\sqrt{3}}{\pi} \int_0^{q_\alpha} \frac{dq}{(q^2+\frac{1}{2})\,(q^2+\frac{2}{3})^{\frac{1}{2}}} = \frac{6}{\pi} \tan^{-1}\left\{ \frac{q_\alpha}{(2+3q_\alpha^2)^{\frac{1}{2}}} \right\},$$

and therefore $$\frac{q_\alpha}{(2+3q_\alpha^2)^{\frac{1}{2}}} = \tan\left\{ \frac{\pi}{6}(1-\alpha) \right\}.$$

If $\tan\left\{\frac{\pi}{6}(1-\alpha)\right\}$ be denoted by τ, then

$$q_\alpha(1,3) = \frac{\sqrt{2}\,\tau}{(1-3\tau^2)^{\frac{1}{2}}}. \tag{20}$$

The six required values of q_α are found by substitution of the corresponding values of α in (20) above, and further multiplication by $d_3 = 3/\sqrt{\pi}$ gives the percentage limits of the distribution of $u = u(1,3)$.

General case $n \geqslant 4$, $m = 1$

For $n \geqslant 4$ no suitable algebraic expression exists for the distribution of the range, but Pearson & Hartley (1942) have tabulated values of the probability integral $\int_0^w p(w)\,dw$ to 4 figures at intervals of 0·05 of w for values of n from 2 to 20. Hartley kindly lent manuscript tables of the integral tabulated to 5 figures at intervals of 0·25 of w for values of n from 2 to 16. Using these five-figure tables for $n = 4, 6, 10, 16$ and the four-figure tables for $n = 20$, the frequency distribution of w was obtained numerically by subtraction of successive values of $\int_0^w p(w)\,dw$ at intervals of 0·25 and then converting these class frequencies into ordinates $y(w)$. The degree of approximation in the formula used implied the vanishing of fifth differences (see K. Pearson, *Tables for Statisticians and Biometricians*, Part II, p. xvii).

Each case was treated in turn, and the six values of the percentage points q_α, corresponding to the six different values of α under consideration, were determined using the relations given in (10) above. Taking a trial value of q_α, the integrals

$$I(w, q_\alpha) = \frac{2}{\sqrt{(2\pi)}}\int_0^{wq_\alpha} e^{-\frac{1}{2}x^2}dx \tag{21}$$

were calculated at intervals of 0·25 over the whole range of w. Quadrature was then applied to the products $y(w)\, I(w, q_\alpha)$ over the range $0 \leqslant w < \infty$, to obtain the value of $(1-\alpha)$ corresponding to the trial value of q_α. This procedure was repeated a number of times to obtain values of $(1-\alpha)$ corresponding to a series of equidistant values of q_α. The required values of q_α corresponding to the six values of α under investigation were then obtained by backward interpolation.

As $n \to \infty$ the ratio w/d_n tends to the population value of the standard deviation. Furthermore, for $n = 2$ and $n = 3$, exact values of q_α had been previously obtained by direct calculation for the six values of α. Thus it was possible to make initial estimates of the required values of q_α, and the process of this 'trial and error' method was not found too laborious.

Table 1. *Framework values of percentage points of $u = u(1, n)$*

n \ α	0·10	0·05	0·02	0·01	0·002	0·001
2	5·0376	10·1381	25·389	50·791	253·97	507·95⁻
3	2·5935⁻	3·8225⁺	6·188	8·819	19·84	28·08
4	2·1793	2·9505⁺	4·213	5·420	9·42	11·75⁺
6	1·9354	2·4755⁺	3·249	3·900	5·71	6·66
10	1·8064	2·2390	2·807	3·244	4·32	4·82
16	1·7496	2·1385⁺	2·628	2·990	3·82	4·19
20 (a)	1·7320	2·1083	2·576	2·916	3·69	4·01
20 (b)	1·7314	2·1074	2·576	2·916	3·69	4·02
∞	1·6449	1·9600	2·326	2·576	3·09	3·29

The framework values of the percentage points of $u(1, n)$ were obtained by multiplying the values of q_α by the corresponding values of the mean range d_n tabulated by Tippett (1925) and are given in Table 1, together with the exact values for $n = 2$ and $n = 3$ determined above.

As a check on the accuracy of determination of these percentage points, the six values for $n = 20$ were also calculated by a second method. Writing

$$r = 1/q = d_n/x, \tag{22}$$

the method is to determine, for the six values of α, the corresponding values of q_α such that

$$\alpha = \int_{-\infty}^{-1/q_\alpha} p(r)\,dr + \int_{1/q_\alpha}^{\infty} p(r)\,dr, \tag{23}$$

where $y = p(r)$ denotes the frequency distribution of r. Since q is distributed symmetrically about zero, its reciprocal r is also distributed symmetrically about zero, and from (22) and (23) it follows that

$$\alpha = 2\int_{1/q_\alpha}^{\infty} p(r)\,dr$$

$$= 2\int_0^{\infty} p(x)\,dx \int_0^{x/q_\alpha} p(w)\,dw. \tag{24}$$

Ordinates of the normal curve $y(x) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}$ were taken at intervals of 0·25 for x in the range $0 \leqslant x < \infty$ from K. Pearson's *Tables for Statisticians and Biometricians*, Part I. Taking a trial value of $1/q_\alpha$, the integrals $\int_0^{x/q_\alpha} p(w)\,dw$ were calculated from Hartley's four-figure tables for each value of x in the above range. By quadrature applied to the products $p(x)\int_0^{x/q_\alpha} p(w)\,dw$ the integral (24) was evaluated. By taking a series of equidistant values of $1/q_\alpha$ other trial values of α were determined. Backward interpolation was then used to obtain the required values of q_α corresponding to the six values of α under consideration. Finally, the six percentage points of u were determined by multiplying the values of q_α by d_{20} given in Tippett's table. These percentage points are given in the penultimate line (b) of Table 1, and comparison with the corresponding values in the line above, (a), indicates good agreement between the two methods of computation.

Since the percentage points of u for $n = 4$, 6, 10 and 16 have been determined using Hartley's five-figure manuscript tables of the cumulative frequency distribution of w, they should be at least as accurate as the percentage points for $n = 20$ determined from the four-figure tables.

Changes in the percentage points at the different levels of significance run most smoothly if arguments proportional to $1/n$ are used in place of n, and reciprocals of u for the variate. Using a six-point general Lagrangian formula applied to the points corresponding to $n = 3, 4, 6, 10, 16$ and 20, values of percentage points of u were determined for $n = 5, 7, 8, 14$ and 18. (In the case of $n = 20$ the mean values of the percentage points determined by the two methods were used.) The interval was then halved, using a nine-point Lagrangian through points corresponding to $n = 4, 6, 8, 10, 12, 14, 16, 18$ and 20. Finally, the six sets of percentage points were differenced as a check, reduced by either one or two figures and, with the exception of those for $n = 5$, are given in the second columns of Tables 3–8 under $m = 1$.

For $n = 5$, the six percentage points of u were independently determined at a later stage of the investigation by the method used above for $n = 4$, 6, 10, 16 and 20, and it is these values which are given in Tables 3–8. In the table below, the values obtained by interpolation from the framework values are compared with those determined by direct calculation.

Percentage points of $u(1, 5)$

$\alpha =$	0·10	0·05	0·02	0·01	0·002	0·001
By direct calculation	2·019	2·635−	3·56	4·38	6·8	8·2
By interpolation	2·020	2·635+	3·56	4·38	6·8	8·2

In one case only, for $\alpha = 0·10$, is there a difference as much as one unit in the last figure, the actual values obtained being 2·0192 by direct computation and 2·0198 by interpolation.

Taking all the various checks into consideration, it appears unlikely that the values of the percentage points of $u(1, n)$ given in the tables at the end of the paper are in error by more than one unit in the last place. The values for $n = 2$ and $n = 3$ are, of course, exact.

(v) COMPUTATION OF PERCENTAGE POINTS OF THE DISTRIBUTION OF $u = u(2, n)$, i.e. CASE WITH $m = 2$

The exact distribution of $u(2, n)$ cannot be determined analytically except in the case of $n = 2$, and hence the various percentage points have necessarily to be evaluated almost wholly by numerical methods.

The probability of the joint occurrence of a pair of ranges w' and w'' from random samples of equal size n from a normal population of unit standard deviation is given by

$$p(w', w'')\,dw'\,dw'' = p(w')\,p(w'')\,dw'\,dw''. \tag{25}$$

If $\bar{w}$ be the mean value of the two ranges, then its distribution is obtained by integrating (25):

$$p(\bar{w})\,d\bar{w} = \int p(w')\,p(w'')\,dw'\,dw'', \tag{26}$$

the integration being taken over the whole field of w' and w'' subject to the conditions

$$\bar{w} = \tfrac{1}{2}(w' + w'') \quad (0 \leqslant w' < \infty,\ 0 \leqslant w'' < \infty). \tag{27}$$

We shall change the variables from w' and w'' to $\bar{w}$ and w', the Jacobian of the transformation being equal to 2. With these new variables, and noting from (27) that w' varies from 0 to $2\bar{w}$, equation (26) gives

$$p(\bar{w}) = 2\int_0^{2\bar{w}} p(w')\,p(2\bar{w} - w')\,dw'. \tag{28}$$

Special case $n = 2$, $m = 2$

In equation (11) above is given the distribution of the range in samples of two random values from a normal population with unit standard deviation. Substituting this in (28) above, the distribution of means of two independent values of w is therefore given by

$$\begin{aligned} p(\bar{w}) &= 2\int_0^{2\bar{w}} \frac{1}{\sqrt{\pi}} e^{-\frac{1}{4}w'^2} \frac{1}{\sqrt{\pi}} e^{-\frac{1}{4}(2\bar{w}-w')^2}\,dw' \\ &= \frac{2}{\pi} e^{-\frac{1}{2}\bar{w}^2} \int_0^{2\bar{w}} e^{-\frac{1}{2}(w'-\bar{w})^2}\,dw' \\ &= \frac{8}{\sqrt{(2\pi)}} e^{-\frac{1}{2}\bar{w}^2} \int_0^{2\bar{w}} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}z^2}\,dz. \end{aligned} \tag{29}$$

Using the notation

$$I(w) = \int_0^{w} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}z^2}\,dz,$$

the expression (29) for the distribution of the mean range may be written in the form

$$p(\bar{w}) = \frac{8}{\sqrt{(2\pi)}} e^{-\frac{1}{2}\bar{w}^2} I(\bar{w}). \tag{30}$$

We may now proceed to determine the distribution of the ratio $q = x/\bar{w}$. Since they are independent, the joint distribution x and $\bar{w}$ is, from (6) and (30), given by

$$p(x, \bar{w})\,dx\,d\bar{w} = \frac{4}{\pi} e^{-\frac{1}{2}x^2} e^{-\frac{1}{2}\bar{w}^2} I(\bar{w})\,dx\,d\bar{w}. \tag{31}$$

Transforming to variables q and $\bar{w}$, noting that the Jacobian of the transformation is equal to $\bar{w}$, and integrating, we obtain

$$p(q) = \frac{4}{\pi}\int_0^\infty e^{-\frac{1}{2}\bar{w}^2(q^2+1)}\, I(\bar{w})\,\bar{w}\,d\bar{w}. \tag{32}$$

Now since

$$\frac{dI(\bar{w})}{d\bar{w}} = \frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}w^2},$$

we make use of the identity

$$\frac{d}{d\bar{w}}\{e^{-\frac{1}{2}\bar{w}^2(q^2+1)}\, I(\bar{w})\} = -w(q^2+1)\,e^{-\frac{1}{2}\bar{w}^2(q^2+1)}\, I(\bar{w}) + \frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}w^2}\,e^{-\frac{1}{2}w^2(q^2+1)}$$

to evaluate the integral in (32) and obtain

$$\begin{aligned} p(q) &= \frac{4}{\pi\sqrt{(2\pi)}\,(q^2+1)}\int_0^\infty e^{-\frac{1}{2}\bar{w}^2}\,e^{-\frac{1}{2}\bar{w}^2(q^2+1)}\,d\bar{w} \\ &= \frac{2}{\pi(q^2+1)\sqrt{(q^2+2)}}. \end{aligned} \tag{33}$$

The ratio q is distributed symmetrically about zero and from (9) and (33) we obtain

$$1-\alpha = \frac{4}{\pi}\int_0^{q_\alpha}\frac{dq}{(q^2+1)\sqrt{(q^2+2)}},$$

and the percentage points are therefore given by

$$q_\alpha = q_\alpha(2,2) = \frac{\tau\sqrt{2}}{1-\tau}, \tag{34}$$

where

$$\tau = \tan\left\{\frac{\pi}{4}(1-\alpha)\right\}.$$

Substitution of $\alpha = 0{\cdot}10$, $0{\cdot}05$, etc., in (34) gives the required values of q_α, and further multiplication by $d_2 = 2/\sqrt{\pi}$ gives the required corresponding percentage points of the distribution of $u(2,2)$.

General case $n \geqslant 3$, $m = 2$

For $n = 4$, 6, 10, 16 and 20, the ordinates $y(w)$, of the distribution of the range have been previously evaluated above at intervals of 0·25 for w, and these are used in place of the unknown $p(w)$. Taking a particular value of $\bar{w}$, quadrature was then applied to the products $y(w)\,y(2\bar{w}-w)$ to obtain a numerical estimate of $p(\bar{w})$ from equation (28). This process was repeated at intervals of 0·25 for $\bar{w}$ through as much of the range $0 \leqslant \bar{w} < \infty$ as was necessary to obtain the required degree of accuracy. For $n = 3$, determination of the ordinates of the distribution of $\bar{w}$ was also based on quadrature, but, in this case, exact figures for the ordinates of the distribution of the range were obtained from its equation found by McKay & Pearson (1933). Because of the rapid rise of the distribution near to the origin, estimates of the lower values of $p(\bar{w})$ in this case were determined using an interval of $\frac{1}{16}$ for $\bar{w}$ in order to obtain the requisite accuracy. For higher values of $\bar{w}$ the interval was progressively decreased to 0·25 used over the tail portion of the curve.

Treating in turn each curve of the distribution of $\bar{w}$ (for $n = 3, 4, 6, 10, 16$ and 20), the values of the percentage points of the distribution of the ratio

$$q = \frac{u}{d_n} = \frac{x}{\bar{w}}$$

were computed by a method similar to that used in evaluating the percentage points of q for the case $m = 1$. Taking trial values of q_α, the integrals

$$I(\bar{w}, q_\alpha) = \frac{2}{\sqrt{(2\pi)}}\int_0^{\bar{w}q_\alpha} e^{-\frac{1}{2}x^2}\,dx$$

were calculated at intervals of 0·25 for $\bar{w}$. Quadrature was then applied to the products $y(\bar{w})\, I(\bar{w}, q_\alpha)$ over the range $0 \leqslant \bar{w} < \infty$ to obtain corresponding values of $(1-\alpha)$, the sum of the two tails of the distribution beyond deviations $\pm q_\alpha$. Repeating this procedure, a series of values of $(1-\alpha)$ were obtained, corresponding to a set of equidistant values of q_α. Backward interpolation was then used to obtain the six values of q_α corresponding to the six values of α under consideration. Finally, the required percentage points of u were obtained

Table 2. *Framework values of percentage points of $u = u(2, n)$*

n \ α	0·10	0·05	0·02	0·01	0·002	0·001
2	2·6203	3·8671	6·266	8·932	20·10	28·47
3	2·0201	2·6365+	3·555+	4·340	6·78	8·66
4	1·8760	2·3672	3·047	3·600	5·07	5·80
6	1·7791	2·1920	2·727	3·133	4·11	4·52
10	1·7225+	2·0926	2·551	2·884	3·64	3·96
16	1·6961	2·0470	2·472	2·775+	3·44	3·71
20	1·6879	2·0329	2·449	2·742	3·38	3·64

by multiplication by the appropriate value of d_n, and are given in Table 2, together with the exact values for $n = 2$ determined from (34).

As before, Lagrangian formulae were used to interpolate the intermediate values of the percentage points of u, again taking arguments proportional to $1/n$ and reciprocals of the percentage points for the variate. As a check, the values were inspected by determining differences up to the third order and then reduced by one place of decimals. With the exception of the values for $n = 5$, the reduced values are given in Tables 3–8.

For $n = 5$ the six percentage points of u were independently determined at a later stage of the investigation by the same methods used for the framework values. These directly computed values are given in the tables mentioned above, and are also reproduced below for comparison with those obtained by interpolation from the framework values.

Percentage points of $u = u(2, 5)$

$\alpha =$	0·10	0·05	0·02	0·01	0·002	0·001
By direct calculation	1·814	2·254	2·84	3·29	4·4	5·0
By interpolation	1·814	2·254	2·84	3·29	4·4	4·9

In the case of $\alpha = 0.001$, direct calculation gives a value of 4·97 compared with 4·92 obtained by interpolation. For other percentage levels the agreement is exact to the number of figures quoted.

(vi) Computation of percentage points of the distribution of $u = u(m, n)$ for $m > 2$

The variance of the mean range in m subgroups of equal size n steadily decreases as m increases, and the ratio $\bar{w}(m,n)/d_n$ gives closer estimates of the population value of the standard deviation of the variate. Hence, following the usual methods of large-sample theory, the limiting values of the percentage points of u, for indefinitely large m, may be determined from integral tables of the normal curves. For a given value of α, the limiting values of the percentage points are, of course, equal for all values of n, and are also equal to the corresponding limiting values of the percentage points of Fisher's t-distribution for an indefinitely large number of degrees of freedom.

In general it was found for a particular value of α and of n that a three-point Lagrangian curve with $1/m$ as argument and reciprocals of the percentage points of u as variate (passing through points corresponding to $m = 1$, 2 and ∞) may be used for interpolation of the required percentage points corresponding to values of m intermediate between 2 and ∞. Only in the case of $n = 2$ and $n = 3$ was the required accuracy not attained by this procedure and further investigation found necessary. Details of the methods used are given below.

In the case of $n = 2$, the percentage points of the distribution of $u(m, 2)$ were also determined for $m = 4$ and $m = 8$ as follows. First, considering $m = 4$, it was necessary to obtain numerical estimates of the ordinates of the distribution of the means of four ranges, each range from a random sample of two values from a normal population with unit standard deviation. Following the method used for $m = 2$ and leading to equation (28), it is easy to show that the distribution of $\bar{w}(2m, n)$, the mean of $2m$ independent ranges each from a random subsample of n values, is given in terms of the distribution of the mean, $\bar{w}(m, n)$, of m such ranges by

$$p(\bar{w}(2m,n)) = 2\int_0^{2\bar{w}(2m,\,n)} p(\bar{w}(m,n))\,p(2\bar{w}(2m,n)-\bar{w}(m,n))\,d\bar{w}(m,n). \tag{35}$$

Using numerical values for $p(\bar{w})$ ($m = 2$, $n = 2$) given by equation (29) above, estimates of the ordinates of the distribution of $p(\bar{w})$ for $m = 4$, $n = 2$ at intervals of 0·25 were found by quadrature methods similar to those described in previous sections by using the above expression. A repetition of this process, using these last computed values, yielded numerical estimates of the distribution of the means of eight ranges, i.e. $m = 8$. Again applying quadrature to the two distributions, values of $(1-\alpha)$ were determined for a series of equidistant values of q_α ($m = 4$ and $m = 8$). The required values of q_α, corresponding to the six values of α between 0·10 and 0·001 under consideration, were then obtained by backward interpolation, and hence the percentage points of $u(4, 2)$ and $u(8, 2)$.

The sets of six percentage points of $u(m, 2)$ were determined for each required value of m by Lagrangian interpolation, reciprocals of m being used as argument and reciprocals of the corresponding percentage points as variate, the curve passing through the points corresponding to $m = 1$, 2, 4, 8 and ∞. The interpolated values of percentage points obtained by this method, and the directly computed values for $m = 4$ and $m = 8$, are given in Tables 3–8

for a series of values of m suitable for practical use. As a check upon the method, the computation was repeated, this time using a four-point Lagrangian passing through points corresponding to $m = 1$, 2, 4 and ∞. Most of these interpolated four-point values of the percentage points agree exactly with the five-point values previously obtained. In cases where differences arise, none exceed 1 unit in the last figure. The five-point Lagrangian method of interpolation therefore certainly appears to be quite adequate for furnishing the required degree of accuracy.

Numerical estimates of the distribution of the means of pairs of ranges from subsamples of size $n = 3$ and $n = 4$ have already been obtained above. Using these in turn in equation (35), estimates of the ordinates of the distributions of the means of four ranges were determined at intervals of 0·25 by quadrature methods. The sets of percentage points of $u(4, 3)$ and $u(4, 4)$ were then computed by the previous method of trial values and subsequent backward interpolation.

For $n = 3$ and $n = 4$, the percentage points of the distribution of u, for given values of m, are lower and nearer to their limiting values than the corresponding points for $n = 2$. Furthermore, the changes in the values of the percentage points for small values of m are also less abrupt. In view of the agreement between the four-point and five-point Lagrangian interpolated values of the percentage points for $n = 2$, a four-point Lagrangian through points corresponding to $m = 1, 2, 4$ and ∞ may certainly be relied upon to give adequate accuracy for the interpolation of percentage points corresponding to intermediate values of m in the case of $n = 3$ and $n = 4$. These values, together with the computed values for $m = 4$, are given in Tables 3–8.

For the remaining values of n, from 5 to 20, the interpolated percentage points given in Tables 3–8 have been obtained by means of a three-point Lagrangian curve, using values of the percentage points corresponding to values of $m = 1$, 2 and ∞. As in the previous cases of interpolation, reciprocals of m and the variate were used in order to obtain small changes in successive differences. To show that this method is adequate, the six sets of percentage points for $n = 4$ were also interpolated using a three-point Lagrangian. In every case except one, these values agreed exactly with the four-point Lagrangian interpolated values previously found and given in Tables 3–8. In the case of the sole exception, the difference between the two interpolated values was only one unit in the last figure. For the less rapidly changing values of the percentage points of u for $n \geqslant 5$, the three-point method of interpolation therefore provides sufficient accuracy for the present purpose.

Taking all checks into consideration it appears that the tabulated values of the percentage points of the distribution of the function $u = u(m, n)$ may be relied upon to the accuracy given: occasionally the values may be one unit in error in the last figure.

In Tables 3 and 4, the 10 % and 5 % points of u were computed to 3 decimal places, but lack of space has necessitated these being curtailed for publication. For the same reason the values of the percentage points of u for the odd values of $n = 11, 13 \ldots 19$ have been omitted. In practical applications of the test, it is not considered that this reduction will cause any undue inconvenience. Fuller tables have, however, been retained for forthcoming work on the power of the u-test and are available for consultation if required.

(vii) APPROXIMATE VALUES OF THE PERCENTAGE POINTS OF u

If there are m subgroups each of n values, and if the estimate of standard deviation is determined as the root-mean-square of the deviations of variate values from the respective

means of the subgroups, then the number of degrees of freedom is $\nu = m(n-1)$. Unlike the usual t-test, when the estimate of standard deviation is determined from the mean range in m subgroups of equal size n, the percentage points of the modified t-distribution investigated above depend upon the relation between m and n. Reference to Tables 3–8 indicates that, for a constant number of degrees of freedom $\nu = m(n-1)$, the values of the percentage points on a given probability level vary slightly as m and n vary. For example, taking $\alpha = 0{\cdot}05$ and $\nu = 8$, we have the following percentage points: 2·272 for $m = 1$ and $n = 9$, 2·254 for $m = 2$ and $n = 5$, 2·250 for $m = 4$ and $n = 3$, and 2·264 for $m = 8$ and $n = 2$. In general, however, the range in the values of the percentage points of u for a given value of ν is small, and this permits the construction of a table giving approximate values of the six sets of percentage points corresponding to different numbers of degrees of freedom.

Approximate values of percentage points of u

Degrees of freedom $\nu = m(n-1)$	Values of α					
	0·10	0·05	0·02	0·01	0·002	0·001
1	5·0	10·1	25·4	50·8	254·0	507·9
2	2·6	3·8	6·2	8·9	19·9	28·3
3	2·2	3·0	4·2	5·5	9·4	11·8
4	2·0	2·6	3·6	4·4	6·8	8·3
5	1·9	2·5	3·3	3·9	5·7	6·7
6	1·9	2·4	3·1	3·6	5·1	5·9
7	1·8	2·3	3·0	3·5	4·7	5·4
8	1·8	2·3	2·9	3·3	4·5	5·0
9	1·8	2·2	2·8	3·2	4·3	4·8
10	1·8	2·2	2·7	3·1	4·2	4·6
11	1·8	2·2	2·7	3·1	4·1	4·5
12	1·8	2·2	2·7	3·1	4·0	4·3
13	1·8	2·2	2·7	3·1	3·9	4·3
14	1·7	2·1	2·6	3·0	3·8	4·2
15	1·7	2·1	2·6	2·9	3·7	4·1
16	1·7	2·1	2·6	2·9	3·7	4·1
17	1·7	2·1	2·6	2·9	3·7	4·1
18	1·7	2·1	2·6	2·9	3·6	4·0
19	1·7	2·1	2·6	2·9	3·6	4·0
20	1·7	2·1	2·5	2·8	3·6	3·9
30	1·7	2·0	2·4	2·7	3·4	3·6
60	1·7	2·0	2·4	2·7	3·2	3·5
120	1·7	2·0	2·4	2·6	3·2	3·4
∞	1·64	1·96	2·33	2·58	3·09	3·29

For a particular pair of values of m and n, the values of the percentage points for $\nu = m(n-1)$ degrees of freedom given in the table above are generally not in error by more than one unit in the last place of figures. This degree of accuracy is frequently sufficient for many practical applications of the distribution of u. To settle the significance of cases giving values of u close to the above approximate values, reference should be made to the accurate values given in Tables 3–8.

(viii) Applications of the u-test

The difference between the mean of a sample of n random values of a normally distributed variate and the population value is shown in the Appendix to be independent of the total

range in the sample, and also independent of the mean range determined from random subgroups of values. The modified t-test based on range estimates of standard deviation may therefore be used in various statistical tests of significance involving deviations of sample means. The application of this range test to sampling problems is analogous to that of the well-known t-test, and no detailed description is therefore required. The most frequent use of the new test will be found in the treatment of experimental data of various types, and also in the examination of test results recorded for the purpose of control of the quality of industrial products. In this latter type of work, cases frequently arise when it is desirable to apply a rapid test for determining the significance of a difference between the mean of a sample and some preassigned value, frequently some desired control level, or the significance of the difference between two sample means. Furthermore, for routine purposes, it is often desirable that the test should not only be rapid but also of a simple nature, thus enabling it to be used by workers with little mathematical or even arithmetical aptitude. The new range test has the advantages of greater simplicity and greatly reduced amount of computing compared with the standard t-test. The use of range estimates of standard deviation, in place of root-mean-square estimates, necessarily entails some loss of precision, but in a future paper it will be shown that this reduction in accuracy is small and certainly negligible for most practical purposes.

The most frequent applications of the range test are considered below and are followed by several numerical examples in which, for purposes of comparison, the parallel treatment by the t-test is also given. As in the t-test, the application of the range test involves the assumption of normality of variate distribution and randomness of sampling. Furthermore, where the standard deviation is estimated from the mean range of several subgroups of values, care should be taken to ensure that the arrangement of these values is also random. This latter condition is usually fulfilled by considering the values in the order in which they were originally recorded. In a few cases, however, the order of recording may not be random; the particular circumstances of a test may be such that the order of the observations may be wholly or partly dependent upon their magnitude. In such cases a set of values can be divided into random subgroups by the use of tables of random sampling numbers or by other means.

(a) *Difference between sample mean and population mean*

Suppose we have some preassigned value ξ, and wish to test whether the mean $\bar{x}$ of a sample of N values may be considered as a reasonable estimate of ξ, or whether the difference between $\bar{x}$ and ξ is real in the statistical sense. The usual assumption, the so-called 'Student's Hypothesis', is made that $\bar{x}$ is the mean of a random sample from a normal population of which the mean is ξ and standard deviation is σ. The differences $(\bar{x}-\xi)$ will be distributed about a mean of zero with a standard error equal to $\sigma/\sqrt{N}$. If the sample be divided into m random subgroups of equal size n, $N = mn$, and $\bar{w}$ is the mean of the m ranges of the subgroups, then the sample estimate of the standard error of the mean is $\bar{w}/(d_n\sqrt{N})$. The ratio of the difference between the means to the estimate of its standard error is

$$u = \frac{|\bar{x}-\xi|\, d_n\sqrt{N}}{\bar{w}}. \tag{36}$$

If the computed value of u exceeds the corresponding percentage point in one of Tables 3–8, then the difference is considered unlikely to have arisen through random sampling on

that particular probability level α. As in the case of the t-test, when considering the asymmetrical case of 'Student's Hypothesis', the values of α at the headings of the tables should be halved.

For fairly small values of N, the estimate of the standard error of the mean may be determined, not from the mean range in subgroups, but from the total range between the maximum and minimum values in the sample. In the notation used above, this corresponds to $m = 1$ and $n = N$. The test of the significance of the difference may be made as above and the computed value of u compared with the percentage points in Tables 3–8. In these cases, however, the computation may be curtailed by using the ratio

$$\frac{\delta}{w} = \frac{u(1,n)}{d_n\sqrt{n}}, \tag{37}$$

where $|\bar{x}-\xi| = \delta$, and w is the range in the undivided sample. Table 9 gives values of the ratio δ/w for various levels of significance corresponding to the sum of the two tails of the distribution. For a chosen level of significance the difference δ is considered too large to have arisen through random sampling errors if the value of δ/w exceeds the corresponding tabulated value. Table 9 will also be found useful for giving a rapid estimate of the accuracy of the mean based on a small number of observations.

(b) *Difference between two sample means*

Suppose the first sample of size N_1 be divided into m_1 random subgroups of size n, and the second sample of size N_2 be divided into m_2 random subgroups also of size n, i.e.

$$n = N_1/m_1 = N_2/m_2.$$

The hypothesis is made that each sample can be considered as a random selection from the same normal population. Let the numerical value of the difference between the two sample means be $|\bar{x}_1-\bar{x}_2|$, and the mean of the (m_1+m_2) ranges of n values be $w = \bar{w}(m_1+m_2, n)$, giving an estimate $\bar{w}/d_n$ for the standard deviation of the variate. The ratio of the difference between the two sample means to the range estimate of the standard error of the difference is

$$u = \frac{|\bar{x}_1-\bar{x}_2|\,d_n}{\bar{w}\sqrt{(1/N_1+1/N_2)}}. \tag{38}$$

The significance of the difference between the means in any particular case can be determined by noting whether the computed value of u exceeds the corresponding percentage point for a chosen value of α by reference to Tables 3–8, using the column headed $m = m_1+m_2$.

When the samples are small and of equal size, say n, the variate standard deviation can be estimated from the two total ranges in the samples. If w' and w'' are the two ranges, with a mean value $\bar{w} = \frac{1}{2}(w'+w'')$, then

$$u = \frac{|\bar{x}_1-\bar{x}_2|\,d_n\sqrt{(\frac{1}{2}n)}}{\bar{w}} \tag{39}$$

may be used as above for testing the significance of the difference between the two means. A more rapid test may, however, be made by simply determining the value of the ratio of the difference between sample means to the average of the two sample ranges

$$\frac{|\bar{x}_1-\bar{x}_2|}{\frac{1}{2}(w'+w'')} = \frac{u(2,n)}{d_n\sqrt{(\frac{1}{2}n)}}. \tag{40}$$

In Table 10 are given values of the above ratio lying on six different probability levels. For

a given level of significance α, values of the ratio smaller than those tabulated may be considered to have arisen through random sampling errors; greater values indicate that a given difference is unlikely to have arisen through chance and therefore point to a real difference.

In the computation of u it is necessary to use values of d_n, the mean range in samples from a normal population of unit standard deviation. A selection of the values determined by Tippett (1925) is reproduced in Table 11 to avoid the necessity of frequent reference to his original paper, and is accompanied by the corresponding values of $\sqrt{n}$ and $d_n\sqrt{n}$.

(c) Confidence intervals

As with 'Student's' test, the tables of percentage points may be used to estimate with a given measure of confidence, the interval within which it can be stated that ξ or $\xi_1 - \xi_2$ lies.

Examples

Example 1. The following data have been previously used as an example by 'Student' (1908). Ten patients were treated with the optical isomers of hyoscyamine hydrobromide and the additional hours of sleep were noted.

Additional hours sleep gained by use of hyoscyamine hydrobromide

Patient	*Dextro*-(D)	*Laevo*-(L)	Difference ($D-L$)
1	+0·7	+1·9	+1·2
2	−1·6	+0·8	+2·4
3	−0·2	+1·1	+1·3
4	−1·2	+0·1	+1·3
5	−0·1	−0·1	0·0
6	+3·4	+4·4	+1·0
7	+3·7	+5·5	+1·8
8	+0·8	+1·6	+0·8
9	0·0	+4·6	+4·6
10	+2·0	+3·4	+1·4
Means	+0·75	+2·33	+1·58

The last column may be used for the controlled comparison of the two drugs, since their effects were measured on the same ten patients. The *laevo* form has given a greater figure for the additional hours sleep than the *dextro* form. Whether the former may be considered as the better soporific is examined by both the standard deviation and range tests.

(*a*) The sum of squares of deviations of the differences about their mean value is 13·616, associated with 9 degrees of freedom. The estimate of the standard error is therefore 0·3890, and the value of t works out to be $1{\cdot}58/0{\cdot}3890 = 4{\cdot}06$. For 9 degrees of freedom a value of $t = 3{\cdot}250$ lies on the 1 % level of significance. Assuming normal random sampling, a value of t equal or greater than 4·06 will occur much less frequently than once in a hundred times. This leads to the conclusion that the *laevo* form is better for producing sleep than the *dextro* form.

(*b*) For examination by the range, the value of $u = u(1, 10)$ may be computed, but in this case it is simpler to use the shortened method of equation (37). The ratio of the mean difference to the range in the ten individual differences is $\delta/w = 1{\cdot}58/4{\cdot}6 = 0{\cdot}34$. Reference to Table 9 shows that this value is slightly in excess of the tabulated value 0·333 on the 1 % level of significance, leading to the same conclusion as that drawn from the t-test.

The greater significance suggested by the t-test seems to be largely due to the exceptional difference $D-L$ for Patient No. 9, viz. 4·6, which affects s more seriously than w.

Example 2. In the calibration of a viscometer it is necessary to time the interval required for the level of an aqueous solution of glycerol to fall between two fixed marks. For satisfactory calibration it is considered desirable that the mean time of flow should be accurate to $\pm \frac{1}{2}$ sec., risking a greater error not more frequently than 1 in 20 times. Five independent determinations of the time interval (in seconds) for one viscometer were 103·5, 104·1, 102·7, 103·2 and 102·6. While this number of observations is clearly too small for a final assessment of accuracy, it is often useful to get an interim answer to guide further action.

(*a*) The sum of squares of the deviations of the five observations about their mean is 1·508, associated with 4 degrees of freedom, giving an estimate of the standard error of the mean equal to 0·275. Reference to tables shows that a value of t equal to 2·776 lies on the 5 % level of significance. Hence in 19 times out of 20 it would be expected that a sample mean will not diverge from the true mean value by more than $\pm 2{\cdot}776 \times 0{\cdot}275 = \pm 0{\cdot}76$ sec. This error exceeds the assigned limits of $\pm \frac{1}{2}$ sec. and therefore points to the necessity of further tests to fulfil the required conditions.

(*b*) Instead of computing an estimate of the standard error of the mean from the range ($w = 104{\cdot}1 - 102{\cdot}6 = 1{\cdot}5$) in the five determinations, we note from Table 9 that a value of $\delta/w = 0{\cdot}507$ lies on the 5 % level of significance. Hence in 19 times out of 20 the sample mean will differ from its true value by an amount up to a deviation of

$$\pm \delta = \pm 0{\cdot}507 \times 1{\cdot}5 = \pm 0{\cdot}76,$$

a result in agreement with that yielded by the t-test.

Example 3. In the processing of raw cotton, modifications were made in the design of one of the machines with the object of improving the efficiency of cleaning. Tests were made on a series of 24 different mixings for the purpose of determining whether yarn strength was adversely affected by the mechanical alterations. The results of the 24 pairs of comparisons are given below (the strength being expressed as a count × strength product), together with the differences between them expressed as percentages of the corresponding strengths under standard conditions.

Yarn strengths under standard and modified conditions

Strength: Standard S	Strength: Modified M	Percentage difference $100(M-S)/S$	Strength: Standard S	Strength: Modified M	Percentage difference $100(M-S)/S$
1805	1763	−2·3	1931	1898	−1·7
1870	1901	+1·7	1508	1520	+0·8
2000	2026	+1·3	2111	2119	+0·4
1823	1904	+4·4	1496	1481	−1·0
1603	1619	+1·0	1672	1723	+3·1
1889	1830	−3·1	1947	1759	−9·7
2058	2019	−1·9	1960	1934	−1·3
1806	1850	+2·4	1624	1594	−1·8
1056	1112	+5·3	2162	2170	+0·4
1857	1782	−4·0	1915	1967	+2·7
1801	1720	−4·5	1738	1810	+4·1
2094	2144	+2·4	1609	1613	+0·2

The mean value for the percentage difference in strength is $-0{\cdot}46$. Whether this is an indication that the mechanical modifications have resulted in the production of weaker yarns is examined by means of the standard deviation and range tests.

(*a*) The sum of squares of the deviations of the percentage differences about their mean value is 256·48, based on 23 degrees of freedom. The estimate of the standard deviation of the percentage differences is 3·34 and the standard error of their mean value is 0·68, giving a value of $t = 0{\cdot}46/0{\cdot}68 = 0{\cdot}59$. This is much below the value of 2·069 on the 5 % level of significance and leads to the conclusion that there are no grounds for suspecting that the mechanical alterations have led to the production of weaker yarns.

(*b*) The number of observations place this case outside the range of Table 9, and it is therefore necessary to use the modified t-function. The data are arranged in random order of their occurrence, and split into four groups of six. The ranges in the sets of six differences are 7·5, 9·8, 12·8 and 5·9 with a mean value $\bar{w}(4, 6) = 9{\cdot}0$. The estimate of the variate standard deviation is $\bar{w}(4, 6)/d_6 = 3{\cdot}55$, giving 0·72 for the standard error of the mean percentage difference, and $u = 0{\cdot}46/0{\cdot}72 = 0{\cdot}64$. The 5 % level of significance is, from Table 4, equal to 2·07, much greater than the value computed from the data and therefore indicates the same conclusion as above.

Example 4. Independent determinations of percentage trash content were made in triplicate on two samples of raw cotton and the following results obtained:

Percentage trash content of raw cotton

	Sample A	Sample B
	1·13	0·76
	1·31	0·64
	1·25	1·01
Means	1·23	0·80

The point to be decided is whether sample B may be said to be cleaner than sample A, or whether the difference between the two average percentage trash contents may be accounted for by random experimental variation. Since, in this case, the comparisons are not paired, the standard error of the difference between the mean values of the two samples has necessarily to be estimated from the variation within each of the two sets of results. As before, normal variation in sampling and in testing errors is assumed.

(*a*) The sum of squares of deviations of each set of values from their mean is 0·01680 for A and 0·07167 for B, each associated with 2 degrees of freedom. The best estimate of the error standard deviation is therefore 0·149, giving 0·122 for the estimate of the standard error of the difference between the two means. The value of t is equal to $(1{\cdot}23-0{\cdot}80)/0{\cdot}122 = 3{\cdot}5$ which exceeds the value 2·776 obtained from tables for 4 degrees of freedom and $\alpha = 0{\cdot}05$. On this level of significance the result is taken to indicate a real difference in the cleanliness of the two cottons.

(*b*) The difference between the two sample means is 0·43 and the mean of the two ranges is 0·275. Hence, using the ratio of equation (40), $|\bar{x}_1-\bar{x}_2|/\frac{1}{2}(w'+w'') = 1{\cdot}6$ which, from Table 10, is seen to exceed the value of 1·272 lying on the 5 % level and therefore is taken to indicate a significant difference in the mean values.

Example 5. The following strength test results were obtained on two batches of cotton yarn (measurements recorded to the nearest $\frac{1}{2}$ lb.) and are noted downwards in order of random occurrence:

Sample A			Sample B	
30·5	31·0	29·5	27·0	28·5
28·0	31·5	27·5	28·5	25·0
29·5	30·0	28·0	26·5	28·0
28·0	27·5	26·0	27·0	27·5
28·5	29·5	28·5	27·0	27·5
29·5	27·5	26·5	28·5	28·5
27·5	28·0	27·0	28·0	28·0
28·0	32·5	30·0	28·0	26·0
29·5	28·5	28·5	25·0	26·5
30·5	29·0	31·0	29·0	28·0

The mean of sample A is 28·90 lb. and 27·40 lb. for sample B, and the question arises as to whether sample B is actually weaker than A.

(a) The sums of squares of the deviations about their respective mean values are 68·2 for A and 24·8 for B, associated with 29 and 19 degrees of freedom. The estimate of the error standard deviation is therefore 1·392, giving 0·402 for the standard error of the difference between the two means and a value of t equal to $(28{\cdot}9-27{\cdot}4)/0{\cdot}402 = 3{\cdot}7$. For 48 degrees of freedom a value of $t = 2{\cdot}68$ lies on the 1 % level of significance. The greater value of 3·7 yielded by the data above indicates, therefore, that the difference in strength of the two yarns may be accepted as statistically significant.

(b) The estimate of the error standard deviation is obtained from the ranges within groups of ten values, three groups for sample A and two for sample B. The values of these five ranges are 3·0, 5·0, 5·0, 4·0 and 3·5 with a mean value of 4·1 and a corresponding estimate of error standard deviation equal to $\bar{w}(5, 10)/d_{10} = 1{\cdot}33$. The estimate of the standard error of the difference between the two means is 0·384 and the value of u is $(28{\cdot}9-27{\cdot}4)/0{\cdot}384 = 3{\cdot}9$. For five ranges of ten, the value of u on the 1 % level of significance is, from Table 6, equal to 2·69 (cf. 2·68 for t with 48 degrees of freedom). The value of 3·9 obtained from the data is greater than this value of 2·69 and this again leads to the conclusion that the difference in mean strengths of the two yarns is 'statistically significant'.

Note added in proof. Since the present paper went to press, a note by Daly (1946) has been published, in which it is suggested that the range may be used in place of the root-mean-square estimate of variance in a test analogous to the t-test. The case where the estimate of standard deviation from a single range is discussed and values of the ratio (deviation)/(range) on the 10 % level of significance are given to two significant figures for a number of low values of n. These agree with the corresponding values given in Table 9, for $\alpha = 0{\cdot}10$, of the present paper. [Mr Lord's paper was first submitted for publication in August 1945. Ed.]

APPENDIX

On the independence of mean and some linear estimates of standard deviation in random samples from a normal population

In the above practical applications of the u distribution to normal random sampling problems, it has been implicitly assumed that range estimates of standard deviation, like root-mean-square estimates, are independent of the mean of the sample from which they have been determined. The validity of this assumption is established below, where it is shown as a particular case of a more general theorem.

Consider a set of n random values of a variable from a normal population of distribution

$$p(x)\,dx = \frac{1}{\sqrt{(2\pi)}\,\sigma}\exp\left[-\frac{1}{2}\frac{(x-\xi)^2}{\sigma^2}\right]dx. \tag{1}$$

Of such a set let x_p and x_q denote the pth and qth values ($p<q$) in ascending order of magnitude, and denote the remaining $(n-2)$ values such that

$$\left.\begin{aligned} -\infty < x_r \leqslant x_p, &\quad r = 1, 2, \ldots, (p-1),\\ x_p \leqslant x_r \leqslant x_q, &\quad r = (p+1), (p+2), \ldots, (q-1),\\ x_q \leqslant x_r < \infty, &\quad r = (q+1), (q+2), \ldots, n. \end{aligned}\right\} \tag{2}$$

Now a set of any n values may be arranged in $n!$ ways and in random samples all arrangements of the same n values are of equal probability. The group of $(p-1)$ values all less than x_p are not ranked in any particular order, and there are hence $(p-1)!$ ways in which they may be arranged. Similarly, the group of values from x_{p+1} to x_{q-1} may be arranged in $(q-p-1)!$ ways and the third group from x_{q+1} to x_n in $(n-q)!$ ways. The distribution of random samples in which the pth and qth values in ascending order are denoted by x_p and x_q, and the remaining values satisfy the conditions in (2), is therefore given by

$$p(x_1, x_2, \ldots, x_n)\,dx_1 dx_2 \ldots dx_n = \left\{\frac{n!}{(p-1)!\,(q-p-1)!\,(n-q)!}\right\}$$
$$\times \frac{1}{(2\pi)^{\frac{1}{2}n}\sigma^n}\exp\left[-\frac{1}{2\sigma^2}\sum_{r=1}^{r=n}(x_r-\xi)^2\right]dx_1 dx_2 \ldots dx_n, \tag{3}$$

where the constant term in brackets makes the total frequency of all such samples equal to unity.

The joint distribution of the sample mean $(\bar{x})$ of the n values and the difference $\Delta = (x_q - x_p)$ is given by

$$p(\bar{x}, \Delta)\,d\bar{x}\,d\Delta = \frac{n!}{(p-1)!\,(q-p-1)!\,(n-q)!}$$
$$\times \frac{1}{(2\pi)^{\frac{1}{2}n}\sigma^n}\int\cdots\int \exp\left[-\frac{1}{2\sigma^2}\sum_{r=1}^{r=n}(x_r-\xi)^2\right]dx_1 dx_2 \ldots dx_n, \tag{4}$$

where the multiple integral is evaluated over the domain of the x's conditioned by the limits indicated in (2) and by $\Delta = (x_q - x_p)$ and $\bar{x} = \frac{1}{n}\sum_{r=1}^{r=n} x_r$.

Make the transformation to variables defined by

$$\left.\begin{aligned}
\bar{x} &= \frac{1}{n}(x_1+x_2+\dots+x_n),\\
y_1 &= -x_p+x_1,\\
y_2 &= -x_p+x_2,\\
&\dots\dots\dots\dots\dots\dots\dots\dots\\
y_{p-1} &= -x_p+x_{p-1},\\
y_{p+1} &= -x_p+x_{p+1},\\
&\dots\dots\dots\dots\dots\dots\dots\dots\\
y_n &= -x_p+x_n.
\end{aligned}\right\} \tag{5}$$

The Jacobian of the transformation is $\dfrac{\partial(x_1\dots x_n)}{\partial(\bar{x},y_1,\dots,y_{p-1},y_{p+1},\dots,y_n)}=1$, and, from (2) and (5), the transformed limits of integration are

$$\left.\begin{aligned}
-\infty<y_r\leqslant 0, &\quad r=1,2,\dots,(p-1),\\
0\leqslant y_r\leqslant \Delta, &\quad r=(p+1),(p+2),\dots,(q-1),\\
\Delta\leqslant y_r<\infty, &\quad r=(q+1),(q+2),\dots,n.\\
y_q=\Delta. &
\end{aligned}\right\} \tag{6}$$

Using the relations in (5) above it may easily be shown that

$$\sum_{r=1}^{r=n}(x_r-\xi)^2 = n(\bar{x}-\xi)^2+\frac{n-1}{n}\sum_{r=1}^{r=n}y_r^2-\frac{2}{n}\sum_{s=1}^{s=n-1}\sum_{t=1}^{t=n-s}y_sy_{s+t}, \tag{7}$$

where in the summation on the right $r,s,s+t\neq p$. With the new variables we have, from (4), (5), (6) and (7), the joint distribution of $\bar{x}$ and Δ given by

$$\begin{aligned}
p(\bar{x},\Delta)\,d\bar{x}\,d\Delta &= \left[\frac{\exp\left[-\dfrac{n(\bar{x}-\xi)^2}{2\sigma^2}\right]}{\sqrt{(2\pi)}\,\sigma/\sqrt{n}}d\bar{x}\right]\times\left[\frac{\sqrt{n}\,(n-1)!\,d\Delta}{(p-1)!\,(q-p-1)!\,(n-q)!\,(2\pi)^{\frac{1}{2}(n-1)}\sigma^{n-1}}\right.\\
&\quad\times\int_{-\infty}^{0}dy_1\dots\int_{-\infty}^{0}dy_{p-1}\int_0^{\Delta}dy_{p+1}\dots\int_0^{\Delta}dy_{q-1}\int_{\Delta}^{\infty}dy_{q+1}\dots\int_{\Delta}^{\infty}\exp\left[-\frac{1}{2n\sigma^2}\right.\\
&\quad\left.\left.\times\left\{(n-1)\sum_{r=1}^{r=n}y_r^2-2\sum_{s=1}^{s=n-1}\sum_{t=1}^{t=n-s}y_sy_{s+t}\right\}\right]dy_n\right].
\end{aligned} \tag{8}$$

with the restriction that $r,s,s+t\neq p$, and Δ is to be substituted for y_q.

The term in the first bracket of (8) is the distribution of the sample mean $\bar{x}$. It follows, therefore, that the term in the second bracket is the distribution of Δ, because this expression does not involve $\bar{x}$ but is a function of Δ alone. This indicates that, in random samples from a normal population, the difference between the pth and qth values in order of magnitude is independent of the sample mean. It follows, therefore, that all estimates of the population standard deviation σ determined from ranked variate differences (e.g. from the semi-interquartile range or other percentile measures of dispersion) are independent of the corresponding sample mean.

As a special case, when $p=1$ and $q=n$, the difference between the pth and qth values becomes the difference between the lowest and highest, i.e. the range of the sample. Furthermore, if the values in a sample be divided into random subgroups, a simple extension of the argument shows that there is also statistical independence between sample mean and the corresponding mean range of the subgroups.

Table 3. 10 % *points of* $u = u(m, n)$

n \ m	1	2	3	4	5	6	8	10	15	20	30	60
2	5·04	2·62	2·20	2·03	1·94	1·89	1·82	1·78	1·73	1·71	1·69	1·67 (1)
3	2·59	2·02	1·88	1·81	1·77	1·75+	1·72	1·71	1·69	1·67	1·66	1·66 (1)
4	2·18	1·88	1·79	1·75+	1·73	1·72	1·70	1·69	1·67	1·67	1·66	1·65+
5	2·02	1·81	1·75+	1·73	1·71	1·70	1·68	1·68	1·67	1·66	1·66	1·65
6	1·94	1·78	1·73	1·71	1·70	1·69	1·68	1·67	1·66	1·66	1·65+	1·65−
7	1·88	1·76	1·72	1·70	1·69	1·68	1·67	1·67	1·66	1·66	1·65+	1·65−
8	1·85	1·74	1·71	1·69	1·68	1·68	1·67	1·66	1·66	1·65+	1·65+	1·65−
9	1·82	1·73	1·70	1·69	1·68	1·67	1·67	1·66	1·66	1·65+	1·65	1·65−
10	1·81	1·72	1·70	1·68	1·68	1·67	1·66	1·66	1·65+	1·65+	1·65	1·65−
12	1·78	1·71	1·69	1·68	1·67	1·67	1·66	1·66	1·65+	1·65+	1·65−	1·65−
14	1·76	1·70	1·68	1·67	1·67	1·66	1·66	1·66	1·65+	1·65	1·65−	1·65−
16	1·75	1·70	1·68	1·67	1·67	1·66	1·66	1·65+	1·65+	1·65	1·65−	1·65−
18	1·74	1·69	1·68	1·67	1·66	1·66	1·66	1·65+	1·65+	1·65−	1·65−	1·65−
20	1·73	1·69	1·67	1·67	1·66	1·66	1·66	1·65+	1·65	1·65−	1·65−	1·65−

Table 4. 5 % *points of* $u = u(m, n)$

n \ m	1	2	3	4	5	6	8	10	15	20	30	60
2	10·14	3·87	2·98	2·66	2·49	2·38	2·26	2·20	2·11	2·07	2·03	2·00 (2)
3	3·82	2·64	2·37	2·25	2·19	2·14	2·09	2·07	2·03	2·01	1·99	1·98 (1)
4	2·95+	2·37	2·22	2·15−	2·11	2·08	2·05	2·03	2·01	2·00	1·98	1·97
5	2·63	2·25+	2·15−	2·10	2·07	2·05	2·03	2·01	2·00	1·99	1·98	1·97 (1)
6	2·48	2·19	2·11	2·07	2·05−	2·03	2·01	2·00	1·99	1·98	1·97	1·97 (1)
7	2·38	2·15+	2·09	2·05+	2·03	2·02	2·01	2·00	1·98	1·98	1·97	1·97 (1)
8	2·32	2·13	2·07	2·04	2·02	2·01	2·00	1·99	1·98	1·98	1·97	1·97 (1)
9	2·27	2·11	2·06	2·03	2·02	2·01	2·00	1·99	1·98	1·97	1·97	1·96
10	2·24	2·09	2·05−	2·02	2·01	2·00	1·99	1·98	1·98	1·97	1·97	1·96
12	2·19	2·07	2·03	2·01	2·00	2·00	1·99	1·98	1·97	1·97	1·97	1·96
14	2·16	2·06	2·02	2·01	2·00	1·99	1·98	1·98	1·97	1·97	1·97	1·96
16	2·14	2·05−	2·02	2·00	1·99	1·99	1·98	1·98	1·97	1·97	1·97	1·96
18	2·12	2·04	2·01	2·00	1·99	1·99	1·98	1·98	1·97	1·97	1·97	1·96
20	2·11	2·03	2·01	2·00	1·99	1·98	1·98	1·97	1·97	1·97	1·96	1·96

Note. The numbers in brackets in the column headed $m = 60$ indicate the number of units which must be subtracted in the second decimal place to obtain the level for $m = 120$ and the same value of n. Where no figure is given $u(120, n) = u(60, n)$ to second decimal place accuracy. E.g. for the 5 % level, $u(120, 2) = 1{\cdot}98$.

Table 5. 2 % *points of* $u = u(m, n)$

n \ m	1	2	3	4	5	6	8	10	15	20	30	60
2	25·39	6·27	4·27	3·60	3·27	3·08	2·86	2·73	2·59	2·52	2·45+	2·39 (3)
3	6·19	3·56	3·05−	2·84	2·72	2·65−	2·56	2·51	2·45	2·42	2·39	2·36 (2)
4	4·21	3·05−	2·77	2·65−	2·58	2·53	2·48	2·45−	2·41	2·39	2·37	2·35− (1)
5	3·56	2·84	2·65−	2·56	2·51	2·48	2·44	2·42	2·39	2·37	2·36	2·34 (1)
6	3·25	2·73	2·58	2·51	2·47	2·45−	2·42	2·40	2·37	2·36	2·35+	2·34 (1)
7	3·07	2·66	2·54	2·48	2·45+	2·43	2·40	2·39	2·37	2·36	2·35−	2·34 (1)
8	2·95+	2·61	2·51	2·46	2·43	2·42	2·39	2·38	2·36	2·35+	2·34	2·34 (1)
9	2·87	2·58	2·49	2·45−	2·42	2·41	2·39	2·37	2·36	2·35+	2·34	2·33
10	2·81	2·55+	2·47	2·44	2·41	2·40	2·38	2·37	2·35+	2·35−	2·34	2·33
12	2·72	2·51	2·45−	2·42	2·40	2·39	2·37	2·36	2·35+	2·34	2·34	2·33
14	2·67	2·49	2·43	2·41	2·39	2·38	2·37	2·36	2·35−	2·34	2·34	2·33
16	2·63	2·47	2·42	2·40	2·38	2·37	2·36	2·35+	2·35−	2·34	2·34	2·33
18	2·60	2·46	2·41	2·39	2·38	2·37	2·36	2·35+	2·34	2·34	2·33	2·33
20	2·58	2·45−	2·41	2·39	2·37	2·37	2·36	2·35−	2·34	2·34	2·33	2·33

Table 6. 1 % *points of* $u = u(m, n)$

n \ m	1	2	3	4	5	6	8	10	15	20	30	60
2	50·79	8·93	5·49	4·43	3·93	3·64	3·32	3·14	2·93	2·84	2·75−	2·66 (4)
3	8·82	4·34	3·60	3·30	3·14	3·03	2·91	2·84	2·75−	2·70	2·66	2·62 (2)
4	5·42	3·60	3·20	3·02	2·92	2·86	2·79	2·74	2·68	2·66	2·63	2·60 (1)
5	4·38	3·29	3·02	2·90	2·83	2·79	2·73	2·70	2·66	2·64	2·62	2·60 (1)
6	3·90	3·13	2·93	2·83	2·78	2·74	2·70	2·67	2·64	2·62	2·61	2·59 (1)
7	3·63	3·03	2·87	2·79	2·75−	2·72	2·68	2·66	2·63	2·62	2·60	2·59 (1)
8	3·45+	2·97	2·83	2·76	2·72	2·70	2·67	2·65−	2·62	2·61	2·60	2·59 (1)
9	3·33	2·92	2·80	2·74	2·71	2·68	2·66	2·64	2·62	2·61	2·60	2·59 (1)
10	3·24	2·88	2·78	2·72	2·69	2·67	2·65	2·63	2·61	2·60	2·59	2·59 (1)
12	3·12	2·83	2·74	2·70	2·68	2·66	2·64	2·62	2·61	2·60	2·59	2·58
14	3·05	2·80	2·72	2·69	2·66	2·65−	2·63	2·62	2·60	2·60	2·59	2·58
16	2·99	2·78	2·71	2·67	2·65+	2·64	2·62	2·61	2·60	2·60	2·59	2·58
18	2·95	2·76	2·70	2·66	2·65−	2·63	2·62	2·61	2·60	2·59	2·59	2·58
20	2·92	2·74	2·69	2·66	2·64	2·63	2·62	2·61	2·60	2·59	2·59	2·58

Note. The numbers in brackets in the column headed $m = 60$ indicate the number of units which must be subtracted in the second decimal place to obtain the level for $m = 120$ and the same value of n. Where no figure is given $u(120, n) = u(60, n)$ to second decimal place accuracy. E.g. for the 2 % level $u(120, 5) = 2{\cdot}33$.

Table 7. 0·2 % *points of* $u = u(m, n)$

n \ m	1	2	3	4	5	6	8	10	15	20	30
2	254·0	20·1	9·4	6·8	5·7	5·1	$4{\cdot}5^-$	4·1	3·7	3·6	3·4 (2)
3	19·8	6·8	5·1	4·5	4·2	4·0	3·7	3·6	3·4	3·3	3·2
4	9·4	5·1	4·2	3·9	3·7	3·6	$3{\cdot}5^-$	3·4	3·3	3·2	3·2 (1)
5	6·8	4·4	3·9	3·7	$3{\cdot}5^+$	$3{\cdot}5^-$	3·4	3·3	3·2	3·2	3·2 (1)
6	5·7	4·1	3·7	$3{\cdot}5^+$	3·4	3·4	3·3	3·3	3·2	3·2	3·1
7	5·1	3·9	3·6	$3{\cdot}5^-$	3·4	3·3	3·3	3·2	3·2	3·2	3·1
8	4·7	3·8	$3{\cdot}5^+$	3·4	3·3	3·3	3·2	3·2	3·2	3·2	3·1
9	4·5	3·7	$3{\cdot}5^-$	3·4	3·3	3·3	3·2	3·2	3·2	3·1	3·1
10	4·3	3·6	3·4	3·3	3·3	3·3	3·2	3·2	3·2	3·1	3·1
12	4·1	$3{\cdot}5^+$	3·4	3·3	3·3	3·2	3·2	3·2	3·1	3·1	3·1
15	3·9	$3{\cdot}5^-$	3·3	3·3	3·2	3·2	3·2	3·2	3·1	3·1	3·1
20	3·7	3·4	3·3	3·2	3·2	3·2	3·2	3·1	3·1	3·1	3·1

Table 8. 0·1 % *points of* $u = u(m, n)$

n \ m	1	2	3	4	5	6	8	10	15	20	30
2	507·9	$28{\cdot}5^-$	11·7	8·1	6·7	5·9	5·0	4·6	4·1	3·9	3·7 (2)
3	28·1	8·7	6·0	5·1	4·6	4·3	4·0	3·9	3·7	3·6	3·5 (1)
4	11·8	5·8	4·7	4·3	4·1	3·9	3·7	3·6	$3{\cdot}5^+$	$3{\cdot}5^-$	3·4 (1)
5	8·2	5·0	4·3	4·0	3·8	3·7	3·6	3·6	$3{\cdot}5^-$	3·4	3·4 (1)
6	6·7	$4{\cdot}5^+$	4·0	3·8	3·7	3·6	$3{\cdot}5^+$	$3{\cdot}5^-$	3·4	3·4	3·4 (1)
7	5·9	4·3	3·9	3·7	3·6	3·6	$3{\cdot}5^+$	$3{\cdot}5^-$	3·4	3·4	3·3
8	5·4	4·1	3·8	3·7	3·6	$3{\cdot}5^+$	$3{\cdot}5^-$	3·4	3·4	3·4	3·3
9	5·0	4·0	3·8	3·6	3·6	$3{\cdot}5^+$	$3{\cdot}5^-$	3·4	3·4	3·4	3·3
10	4·8	4·0	3·7	3·6	$3{\cdot}5^+$	3·5	3·4	3·4	3·4	3·4	3·3
12	$4{\cdot}5^+$	3·8	3·6	3·6	3·5	$3{\cdot}5^-$	3·4	3·4	3·4	3·3	3·3
15	4·2	3·7	3·6	$3{\cdot}5^+$	$3{\cdot}5^-$	3·4	3·4	3·4	3·3	3·3	3·3
20	4·0	3·6	$3{\cdot}5^+$	$3{\cdot}5^-$	3·4	3·4	3·4	3·4	3·3	3·3	3·3

Note. The numbers in brackets in the column headed $m = 30$ indicate the number of units which must be subtracted in the first decimal place to obtain the level for $m = 60$ and the same value of n. Where no figure is given $u(60, n) = u(30, n)$ to first decimal place accuracy; $u(120, n) = u(60, n)$ for (i) all 0·2 % points except that $u(120, 3) = 3{\cdot}1$ and (ii) all 0·1 % points except that $u(120, 2) = 3{\cdot}4$ and $u(120, 3) = 3{\cdot}3$.

Table 9. *Table for testing the significance of the deviation of the mean of a small sample (of size n) from some pre-assigned value*

n \ α	0·10	0·05	0·02	0·01	0·002	0·001
2	3·196	6·353	15·910	31·828	159·16	318·31
3	0·885−	1·304	2·111	3·008	6·77	9·58
4	·529	0·717	1·023	1·316	2·29	2·85+
5	·388	·507	0·685+	0·843	1·32	1·58
6	0·312	0·399	0·523	0·628	0·92	1·07
7	·263	·333	·429	·507	·71	0·82
8	·230	·288	·366	·429	·59	·67
9	·205−	·255+	·322	·374	·50	·57
10	·186	·230	·288	·333	·44	·50
11	0·170	0·210	0·262	0·302	0·40	0·44
12	·158	·194	·241	·277	·36	·40
13	·147	·181	·224	·256	·33	·37
14	·138	·170	·209	·239	·31	·34
15	·131	·160	·197	·224	·29	·32
16	0·124	0·151	0·186	0·212	0·27	0·30
17	·118	·144	·177	·201	·26	·28
18	·113	·137	·168	·191	·24	·26
19	·108	·131	·161	·182	·23	·25+
20	·104	·126	·154	·175−	·22	·24

The table gives values of the ratio $\dfrac{\delta}{w} = \dfrac{\text{deviation of sample mean}}{\text{range in sample}}$ lying on different levels of significance, the levels being the sum, α, of the two tails of the probability distribution.

Table 10. *Table for testing the significance of the difference between the means of two small samples of equal size n*

n \ α	0·10	0·05	0·02	0·01	0·002	0·001
2	2·322	3·427	5·553	7·916	17·81	25·23
3	0·974	1·272	1·715−	2·093	3·27	4·18
4	·644	0·813	1·047	1·237	1·74	1·99
5	·493	·613	0·772	0·896	1·21	1·35+
6	0·405+	0·499	0·621	0·714	0·94	1·03
7	·347	·426	·525+	·600	·77	0·85−
8	·306	·373	·459	·521	·67	·73
9	·275−	·334	·409	·464	·59	·64
10	·250	·304	·371	·419	·53	·58
11	0·233	0·280	0·340	0·384	0·48	0·52
12	·214	·260	·315+	·355+	·44	·48
13	·201	·243	·294	·331	·41	·45−
14	·189	·228	·276	·311	·39	·42
15	·179	·216	·261	·293	·36	·39
16	0·170	0·205−	0·247	0·278	0·34	0·37
17	·162	·195+	·236	·264	·33	·35+
18	·155+	·187	·225+	·252	·31	·34
19	·149	·179	·216	·242	·30	·32
20	·143	·172	·207	·232	·29	·31

The table gives values of the ratio $\dfrac{|\bar{x}_1 - \bar{x}_2|}{\frac{1}{2}(w' + w'')} = \dfrac{\text{difference between means}}{\text{mean of sample ranges}}$ lying on different levels of significance. The levels are the sum, α, of the two tails of the probability distribution.

N.B. When considering deviations in the positive (or negative) direction only, the values of α at the headings of the columns should be halved.

Table 11

n	d_n	$1/d_n$	$\sqrt{n}$	$d_n\sqrt{n}$
2	1·1284	0·8862	1·4142	1·5958
3	1·6926	·5908	1·7321	2·9316
4	2·0588	·4857	2·0000	4·1175
5	2·3259	·4299	2·2361	5·2009
6	2·5344	0·3946	2·4495	6·2080
7	2·7044	·3698	2·6458	7·1551
8	2·8472	·3512	2·8284	8·0531
9	2·9700	·3367	3·0000	8·9101
10	3·0775	·3249	3·1623	9·7319
11	3·1729	0·3152	3·3166	10·5232
12	3·2585	·3069	3·4641	11·2876
13	3·3360	·2998	3·6056	12·0281
14	3·4068	·2935	3·7417	12·7469
15	3·4718	·2880	3·8730	13·4463
16	3·5320	0·2831	4·0000	14·1279
17	3·5879	·2787	4·1231	14·7932
18	3·6401	·2747	4·2426	15·4435
19	3·6890	·2711	4·3589	16·0798
20	3·7350	·2677	4·4721	16·7032

REFERENCES

DALY, J. F. (1946). *Ann. Math. Statist.* **17**, 71.

DAVIES, O. L. & PEARSON, E. S. (1934). *J. Roy. Statist. Soc. Suppl.* **1**, 76.

FISHER, R. A. (1925). *Metron*, **5**, 90.

HARTLEY, H. O. (1942). *Biometrika*, **32**, 334.

MCKAY, A. T. & PEARSON, E. S. (1933). *Biometrika*, **25**, 415.

PEARSON, E. S. (1926). *Biometrika*, **18**, 173.

PEARSON, E. S. (1932). *Biometrika*, **24**, 404.

PEARSON, E. S. (1935). *The Application of Statistical Methods to Industrial Standardization and Quality Control.* B.S. No. 600. London: British Standards Institution.

PEARSON, E. S. & HAINES, JOAN (1935). *J. Roy. Statist. Soc. Suppl.* **2**, 83.

PEARSON, E. S. & HARTLEY, H. O. (1942). *Biometrika*, **32**, 301.

'STUDENT' (1908). *Biometrika*, **6**, 1.

TIPPETT, L. H. C. (1925). *Biometrika*, **17**, 364.

THE FREQUENCY DISTRIBUTION OF $\sqrt{b_1}$ FOR SAMPLES OF ALL SIZES DRAWN AT RANDOM FROM A NORMAL POPULATION

By R. C. GEARY

1. Introductory

A research on which the writer has been engaged for some years has so far yielded the following results:

(1) Testing for normality has a greater practical importance than statisticians (including the writer) have been disposed to accord to it; actual probabilities may be seriously at variance with probabilities derived from the well-known tables computed on the hypothesis of universal normality; in consequence, testing for normality and, where necessary, correction (even if rough and tentative) for suspected universal non-normality, should become a part of statistical routine.

(2) For large samples, $\sqrt{b_1}$ and b_2 are the most efficient of large fields of tests of skewness and kurtosis, respectively, amongst large fields of alternative universes.

These matters will be dealt with in detail in subsequent papers. It seems, in the first instance, desirable to derive the frequency distribution of $\sqrt{b_1}$ for normal random samples of all sizes, partly on account of the inherent importance of the problem, partly in order to explore a computational technique which might be found effective in solving the analogous but probably more difficult b_2 problem.

Towards the solution of the problem there are available the exact values of first four even moments—the odd moments are, of course, zero—of normal $\sqrt{b_1}$, the second, fourth and sixth having been determined by R. A. Fisher (1930) and the eighth by Joseph Pepper (1932). It may be useful here to set out the four moments. Taking

$$\sqrt{b_1} = m_3/m_2^{\frac{3}{2}} = n^{\frac{1}{2}}\left\{\sum_{i=1}^{n}(x_i-\bar{x})^3\right\}\Big/\{\Sigma(x_i-\bar{x})^2\}^{\frac{3}{2}}, \tag{1·1}$$

where n is the sample number, we have

$$\left.\begin{aligned}
\mu_2 &= \frac{6(n-2)}{(n+1)(n+3)},\\
\mu_4 &= \frac{108(n-2)(n^2+27n-70)}{(n+1)(n+3)(n+5)(n+7)(n+9)},\\
\mu_6 &= \frac{3240(n-2)(n^4+84n^3+2695n^2-15168n+20020)}{(n+1)(n+3)(n+5)(n+7)(n+9)(n+11)(n+13)(n+15)},\\
\mu_8 &= \frac{7.5.3^5.2^4(n-2)(n^6+171n^5+13893n^4+580401n^3-5131014n^2+14132268n-12932920)}{(n+1)(n+3)(n+5)\ldots(n+17)(n+19)(n+21)}.
\end{aligned}\right\} \tag{1·2}$$

E. S. Pearson (1931, 1936) derived empirically 0·05 and 0·01 probability points for certain values of $n \geqslant 25$ using a Pearson Type VII curve and earlier approximations by R. A. Fisher (1929) of the second and fourth moments.

The method here used for the derivation of the frequency distribution of $\sqrt{b_1}$ is essentially an elaboration of that which the author used (1935, 1936) for finding the frequency distribution of the test of kurtosis a (the ratio of the mean deviation to the standard deviation of the numbers sampled), which consisted in establishing a relation in integral form between the frequency ordinate for n with the value for $(n-1)$ and thereby determining the ordinates to any required degree of accuracy for the lower n's. At a certain stage the actual frequency is shown to be very close to the value based on the Gram-Charlier curve for the same value of n; and the assumption is made that the Gram-Charlier may be relied on for values of n greater than the 'transition value'. In the present problem the known normal moments are utilized as well at every stage. In the concluding section the status of the solution in the hierarchy of 'precision' is discussed.

Since the frequency is symmetrical, attention is confined practically exclusively to the positive sector.

2. THE GENERAL INTEGRAL ITERATION

To distinguish the sample size by the notation let the value of $\sqrt{b_1}$ be indicated by t_n. Apply a Helmert orthogonal transformation to the original observations $x_1, x_2, \ldots, x_n$ so that

$$\left.\begin{aligned} x_1' &= (x_1 - x_2)/\sqrt{2}, \\ x_2' &= (x_1 + x_2 - 2x_3)/\sqrt{6}, \\ &\vdots \\ x_{n-1}' &= (x_1 + x_2 + \ldots + x_{n-1} - \overline{n-1}\,x_n)/\sqrt{[n(n-1)]}, \\ x_n' &= (x_1 + x_2 + \ldots + x_n)/\sqrt{n} = \bar{x}\sqrt{n}, \end{aligned}\right\} \tag{2·1}$$

which, on inversion, gives

$$\left.\begin{aligned} x_1 - \bar{x} &= \frac{x_1'}{\sqrt{2}} + \frac{x_2'}{\sqrt{6}} + \ldots + \frac{x_{n-1}'}{\sqrt{[n(n-1)]}}, \\ x_2 - \bar{x} &= -\frac{x_1'}{\sqrt{2}} + \frac{x_2'}{\sqrt{6}} + \ldots + \frac{x_{n-1}'}{\sqrt{[n(n-1)]}}, \\ x_3 - \bar{x} &= -\frac{2x_2'}{\sqrt{6}} + \ldots + \frac{x_{n-1}'}{\sqrt{[n(n-1)]}}, \\ &\vdots \\ x_{n-1} - \bar{x} &= -\frac{(n-2)\,x_{n-2}'}{\sqrt{[(n-1)(n-2)]}} + \frac{x_{n-1}'}{\sqrt{[n(n-1)]}}, \\ x_n - \bar{x} &= -\frac{(n-1)\,x_{n-1}'}{\sqrt{[n(n-1)]}}. \end{aligned}\right\} \tag{2·2}$$

Then
$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n-1} x_i'^2, \tag{2·3}$$

and

$$\left.\begin{aligned}\Sigma(x_i-\bar{x})^3 = {} & 3x_1'^2\left(\frac{x_2'}{\sqrt{6}}+\frac{x_3'}{\sqrt{12}}+\dots+\frac{x_{n-1}'}{\sqrt{[(n-1)n]}}\right)\\ & +3x_2'^2\left(\frac{x_3'}{\sqrt{12}}+\frac{x_4'}{\sqrt{20}}+\dots+\frac{x_{n-1}'}{\sqrt{[(n-1)n]}}\right)\\ & +3x_3'^2\left(\frac{x_4'}{\sqrt{20}}+\frac{x_5'}{\sqrt{30}}+\dots+\frac{x_{n-1}'}{\sqrt{[(n-1)n]}}\right)\\ & \vdots\\ & +\frac{3x_{n-2}'^2x_{n-1}'}{\sqrt{[(n-1)n]}}\\ & -\frac{x_2'^3}{\sqrt{6}}-\frac{2x_3'^3}{\sqrt{12}}-\frac{3x_4'^3}{\sqrt{20}}-\dots-\frac{(n-2)x_{n-1}'^3}{\sqrt{[(n-1)n]}}.\end{aligned}\right\} \tag{2·4}$$

Apply a polar transformation to the x_i', that is,

$$\left.\begin{aligned}x_1' &= r\sin\phi_{n-3}\sin\phi_{n-4}\dots\sin\phi_1\sin\phi_0,\\ x_2' &= r\sin\phi_{n-3}\sin\phi_{n-4}\dots\sin\phi_1\cos\phi_0,\\ x_3' &= r\sin\phi_{n-3}\sin\phi_{n-4}\dots\sin\phi_2\cos\phi_1,\\ x_4' &= r\sin\phi_{n-3}\sin\phi_{n-4}\dots\sin\phi_3\cos\phi_2,\\ &\vdots\\ x_{n-3}' &= r\sin\phi_{n-3}\sin\phi_{n-4}\cos\phi_{n-5},\\ x_{n-2}' &= r\sin\phi_{n-3}\cos\phi_{n-4},\\ x_{n-1}' &= r\cos\phi_{n-3},\end{aligned}\right\} \tag{2·5}$$

$$\Sigma x_i'^2 = \Sigma(x_i-\bar{x})^2 = r^2, \tag{2·6}$$

and

$$\begin{aligned}t_n = n^{\frac{1}{2}}\Bigg\{ & \frac{3}{\sqrt{6}}\sin^3\phi_{n-3}\sin^3\phi_{n-4}\dots\sin^3\phi_1\sin^2\phi_0\cos\phi_0\\ & +\frac{3}{\sqrt{12}}\sin^3\phi_{n-3}\sin^3\phi_{n-4}\dots\sin^3\phi_2\sin^2\phi_1\cos\phi_1\\ & +\dots+\frac{3}{\sqrt{[(n-2)(n-1)]}}\sin^3\phi_{n-3}\sin^2\phi_{n-4}\cos\phi_{n-4}+\frac{3}{\sqrt{[(n-1)n]}}\sin^2\phi_{n-3}\cos\phi_{n-3}\\ & -\frac{1}{\sqrt{6}}\sin^3\phi_{n-3}\sin^3\phi_{n-4}\dots\cos^3\phi_0-\frac{2}{\sqrt{12}}\sin^3\phi_{n-3}\dots\sin^3\phi_2\cos^3\phi_1\\ & -\dots-\frac{(n-3)}{\sqrt{[(n-2)(n-1)]}}\sin^3\phi_{n-3}\cos^3\phi_{n-4}-\frac{(n-2)}{\sqrt{[(n-1)n]}}\cos^3\phi_{n-3}\Bigg\},\end{aligned} \tag{2·7}$$

whence the fundamental iteration

$$\frac{t_n}{n^{\frac{1}{2}}} = \frac{t_{n-1}}{(n-1)^{\frac{1}{2}}}\sin^3\phi_{n-3}+\frac{3}{[(n-1)n]^{\frac{1}{2}}}\sin^2\phi_{n-3}\cos\phi_{n-3}-\frac{(n-2)}{[(n-1)n]^{\frac{1}{2}}}\cos^3\phi_{n-3}, \tag{2·8}$$

in which there intervenes only the angle ϕ_{n-3}; and for normal random samples it is a well-known fact that the ϕ_i are distributed independently of one another, the distribution of ϕ_{n-3} being of the form

$$C\sin^{n-3}\phi_{n-3}d\phi_{n-3}. \tag{2·9}$$

Now t_{n-1} involves only $\phi_0, \dots, \phi_{n-4}$; hence it is independent of ϕ_{n-3}. Accordingly, if the frequency distribution of t_{n-1} is of the form

$$f_{n-1}(t_{n-1})\,dt_{n-1}, \tag{2·10}$$

the joint distribution of ϕ_{n-3} and t_{n-1} is given by

$$C \sin^{n-3}\phi_{n-3}\, d\phi_{n-3} \times f_{n-1}(t_{n-1})\, dt_{n-1}. \tag{2·11}$$

Now, from (2·8),

$$dt_{n-1} = dt_n \left(\frac{n-1}{n}\right)^{\frac{1}{2}} \sin^{-3}\phi_{n-3}. \tag{2·12}$$

On substituting in (2·11) and integrating we find for frequency of t_n the expression

$$f_n(t_n) = \left(\frac{n-1}{\pi n}\right)^{\frac{1}{2}} \frac{\frac{1}{2}(n-3)!}{\frac{1}{2}(n-4)!} \int d\phi_{n-3} \sin^{n-6}\phi_{n-3} f_{n-1}(t_{n-1}), \tag{2·13}$$

where the relation (2·8) obtains. Integration extends to values of ϕ_{n-3} (so that $0 \leqslant \phi_{n-3} \leqslant \pi$ for $n > 3$) which yield non-zero values of f_{n-1}. Setting $\cos\phi_{n-3} = x$ the integral at (2·13) assumes the form

$$f_n(t_n) = \left(\frac{n-1}{\pi n}\right)^{\frac{1}{2}} \frac{\frac{1}{2}(n-3)!}{\frac{1}{2}(n-4)!} \int dx (1-x^2)^{\frac{1}{2}(n-7)} f_{n-1}(t_{n-1}), \tag{2·14}$$

with, from (2·8),

$$n^{\frac{1}{2}} t_{n-1} = [(n-1)^{\frac{1}{2}} t_n - 3x + (n+1)x^3]/(1-x^2)^{\frac{3}{2}}. \tag{2·15}$$

In the derivation of the frequencies for $n = 4$ to 8 inclusive, dealt with in later sections, both the forms (2·13) and (2·14) are used.

3. Functional discontinuities of the frequency

In the integral at (2·14) t_n appears merely as a parameter. Consequently the nature of the frequency $f_n(t_n)$ depends to a considerable extent on the simple algebraic properties of $t_{n-1}(x)$ given by (2·15). The following property (easily demonstrated) is fundamental:

For

$$t_n = (n-2k)/[k(n-k)]^{\frac{1}{2}} = {}_k\tau_n \quad (k = 1, 2, \ldots), \tag{3·1}$$

$t_{n-1}(x)$ has a maximum value of

$$(n-2k+1)/[(k-1)(n-k)]^{\frac{1}{2}} = {}_{k-1}\tau_{n-1} \tag{3·2}$$

for

$$x = -[(n-k)/k(n-1)]^{\frac{1}{2}} = {}_k\xi'_n, \tag{3·3}$$

and a minimum value of

$$(n-2k-1)/[k(n-k-1)]^{\frac{1}{2}} = {}_k\tau_{n-1} \tag{3·4}$$

for

$$x = [k/(n-1)(n-k)]^{\frac{1}{2}} = {}_k\xi''_n. \tag{3·5}$$

DEFINITION. ${}_k\tau_n$ are termed the *link values* or *links* of t_n. The regions between consecutive links are termed *zones*. The graph of $t_{n-1}(x)$ for $-1 \leqslant x \leqslant +1$ and $t_n = {}_k\tau_n$ (given at (3·1)) is illustrated in Fig. 1. The limits of integration for integral (2·14) are now seen to be ${}_k\lambda'_n$ and ${}_k\lambda''_n$ which are the values of x at which the ordinates of the curve (2·15) in (x, t_{n-1}), with parameter $t_n = {}_k\tau_n$, assume the limiting values $+{}_1\tau_{n-1}$ and $-{}_1\tau_{n-1}$. The scale on the right shows the links of t_{n-1}. The curve $t_n = {}_k\tau_n$ traverses all the zones but has a 'turn' in the $(k-1)$th zone, remaining entirely in the zone the while. It is due to this turn that the phenomenon of functional discontinuity manifests itself in the frequency $f_n(t_n)$.

Assume that within the kth zone the frequency $f_{n-1}(t_{n-1})$ is represented by ${}_kf_{n-1}(t_{n-1})$, different in functional form for different values of k but the same (for example, having the same coefficients in a power series) within each zone. It will at once be evident, from (2·14), that the frequency of t_n will have a like property. Now, from (2·4) and (2·5) it will be seen that

$$t_3 = -\frac{\cos 3\phi_0}{\sqrt{2}}; \tag{3·6}$$

the distribution of ϕ_0 is rectangular, so that the distribution* of t_3 is given by

$$f_3(t_3) = \frac{\sqrt{2}}{\pi\sqrt{(1-2t_3^2)}}, \quad |t_3| \leqslant 1/\sqrt{2} \tag{3·7}$$

and zero for $|t_3| > 1/\sqrt{2}$. It follows that t_3 has a functional discontinuity at its links $\pm 1/\sqrt{2}$. *Hence, by iteration, the frequency of* t_n *is represented by different functional expressions in its different interlink zones.*

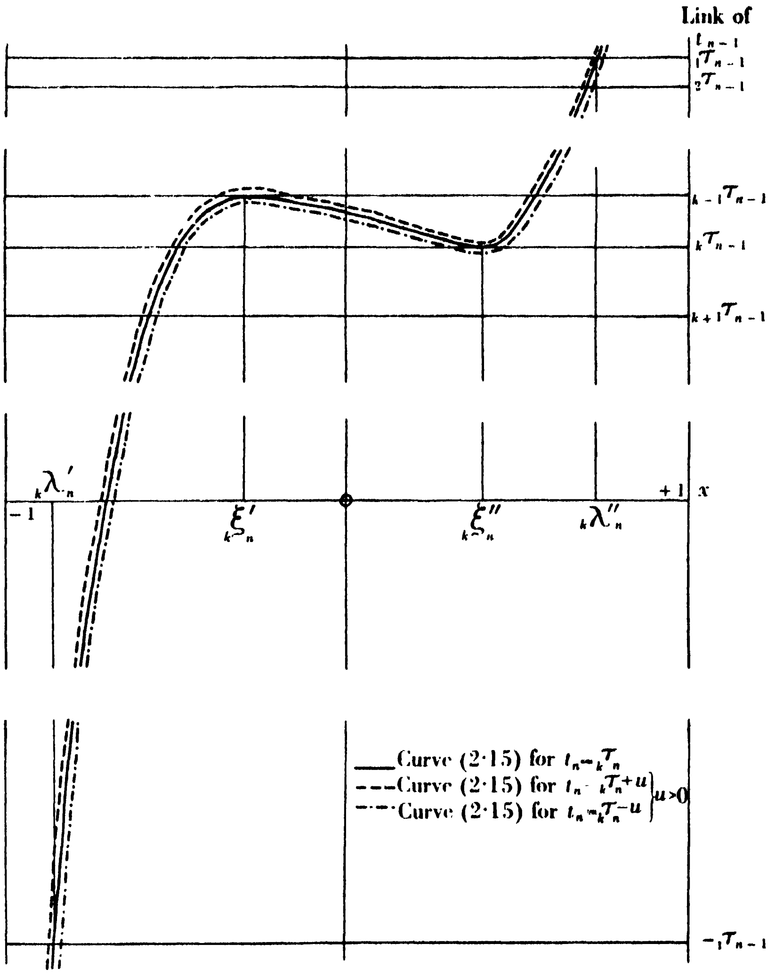


Fig. 1. Graph of $t_{n-1}(x)$.

That the frequency has a finite limit ${}_1\tau_n$ (when n is finite) is established as follows. It can easily be seen from (2·15) that when $t_n = {}_1\tau_n$ the curve $t_{n-1} = t_{n-1}(x)$ degenerates into (i) the straight line $x = -1$ and (ii) a section above the straight line $t_{n-1} = {}_1\tau_{n-1}$ but touching it. For $t_n > {}_1\tau_n$ no part of the curve $t_{n-1} = t_{n-1}(x)$ falls within the rectangle $x = \pm 1, t_{n-1} = \pm {}_1\tau_{n-1}$. Reference to (2·14) shows at once that, if $f_{n-1}(t_{n-1}) = 0$ for $|t_{n-1}| > {}_1\tau_{n-1}$, then $f_n(t_n) = 0$ for $|t_n| > {}_1\tau_n$. But (3·7) shows that t_3 has as limiting values $\pm 1/\sqrt{2}$. Hence, by iteration, it follows that the limiting values of the frequency of t_n (or *simpliciter* of t_n) are

$$\pm\, {}_1\tau_{n-1} = \pm (n-2)/(n-1)^{\frac{1}{2}}. \tag{3·8}$$

* R. A. Fisher (1930).

As will presently appear, the frequencies for $n = 4$ and 5 have marked irregularities: successive integration in accordance with (2·14) imparts, of course, a progressively increasing degree of smoothness to the frequency. To give mathematical expression to this feature, recourse is had to the idea of *order of contact*.

DEFINITION. *Two functions are said to have contact of order* ${}_k\gamma_n$ *at link* ${}_k\tau_n$ *if the functions and their first* $({}_k\gamma_n - 1)$ *derivatives are finite and equal at the link.* It can be shown without difficulty that

$$ {}_k\gamma_n = {}_{k-1}\gamma_{n-1} + 1, \tag{3·9}$$

when $k > 1$, $n > 4$. For what follows it will be convenient to set out for the smaller sample numbers the values of the links and their orders of contact. The links for positive values only of the variables are shown. The orders of contact ${}_1\gamma_n$ will appear from a proposition proved in § 5, giving the actual values of the frequencies near the limit of range. The non-diminishing smoothness in the direction of the centre of the range will be noted.

Values of ${}_k\tau_n$ *and* ${}_k\gamma_n$ *for* $n = 3$ *to* 8 *inclusive*

n	1st link		2nd link		3rd link		4th link	
	${}_1\tau_n$	${}_1\gamma_n$	${}_2\tau_n$	${}_2\gamma_n$	${}_3\tau_n$	${}_3\gamma_n$	${}_4\tau_n$	${}_4\gamma_n$
3	$1/\sqrt{2}$	0	—	—	—	—	—	—
4	$2/\sqrt{3}$	0	0	0	—	—	—	—
5	$3/2$	1	$1/\sqrt{6}$	1	—	—	—	—
6	$4/\sqrt{5}$	1	$1/\sqrt{2}$	2	0	2	—	—
7	$5/\sqrt{6}$	2	$3/\sqrt{10}$	2	$1/2\sqrt{3}$	3	—	—
8	$6/\sqrt{7}$	2	$2/\sqrt{3}$	3	$2/\sqrt{15}$	3	0	4

For even values of n the origin is always a link. In the determination of the frequencies for $n = 5$ to 8, by the methods described in subsequent sections, the link ordinates and the central ordinate play a cardinal role. In fact, the method will be seen to consist essentially in finding curves which pass through the central and link ordinal points, have the required orders of contact and the required form at the limit of range and have the exact earlier momental values (see first section).

4. THE FREQUENCY NEAR THE CENTRE OF RANGE

It will first be shown that

$$f_n'(+0) = 0 \quad \text{for} \quad n > 4. \tag{4·1}$$

In fact, from (2·14) and (2·15) if $t_n = u$, a small positive quantity,

$$f_n(u) = C\int_{-\lambda-\kappa u=\lambda'}^{+\lambda-\kappa u=\lambda''} dx(1-x^2)^{\frac{1}{2}(n-7)} f_{n-1}[t_{n-1}(x)],$$

λ, κ being positive constants. Hence

$$f_n'(u) = -C\kappa\{(1-\lambda''^2)^{\frac{1}{2}(n-7)} f_{n-1}[t_{n-1}(\lambda'')] - (1-\lambda'^2)^{\frac{1}{2}(n-7)} f_{n-1}[t_{n-1}(\lambda')]\}$$

$$+ C'\int_{\lambda'}^{\lambda''} dx(1-x^2)^{\frac{1}{2}(n-10)} f_{n-1}'[t_{n-1}(x)].$$

Letting $u \to 0$ the integral-free expression obviously vanishes provided that $f_{n-1}[t_{n-1}(\lambda)]$ is finite, which it is when $n > 4$; and the integral becomes

$$\int_{-\lambda}^{+\lambda} dx(1-x^2)^{\frac{1}{2}(n-10)} f'_{n-1}\left(\frac{\sqrt{n+1}\, x^3 - 3x}{\overline{1-x^2}^{\frac{3}{2}}}\right).$$

Since $f_{n-1}(y)$ is an even function of y, its derivative is odd which remains an odd function when y is replaced by an odd function of x. Hence the integral vanishes.

5. The form of the frequency at the limit of range

In this section it will be shown that *near* $t_n = \pm(n-2)/(n-1)^{\frac{1}{2}}$ *the frequency is given by*

$$f_n(t_n) = \frac{1}{3\sqrt{\pi}} \frac{\frac{1}{2}(n-3)!}{\frac{1}{2}(n-4)!} \frac{(n-1)^{\frac{1}{2}(n-3)}}{(3n.\overline{n-2})^{\frac{1}{2}(n-4)}} \left(\frac{\overline{n-2}^2}{n-1} - t_n^2\right)^{\frac{1}{2}(n-4)}. \tag{5·1}$$

It may be seen at once that for $n = 3$ the frequency by (5·1) would be

$$f_3(t_3) = \frac{1}{\pi}(\tfrac{1}{2} - t_3^2)^{-\frac{1}{2}}, \tag{5·2}$$

as at (3·7). For $n = 4$, (5·1) gives $\frac{1}{2}\sqrt{3}$, which is the value found by A. T. McKay (1933).

The general theorem will be proved by iteration. We assume a general form

$$f_{n-1}(t_{n-1}) = C_{n-1}\left(\frac{\overline{n-3}^2}{n-2} - t_{n-1}^2\right)^{\frac{1}{2}(n-5)}, \tag{5·3}$$

and show that a similar form emerges for $f_n(t_n)$, finding incidentally an iteration relation for the constant C_n. First set

$$v = \overline{n-2} - \overline{n-1}^{\frac{1}{2}}\, t_n,$$

and assume that v is a positive quantity. It will readily appear, from (2·15), that, for $v = 0$, $t_{n-1}(x)$ has a double root at $x = 1/(n-1)$. Accordingly we set

$$x = x' + 1/(n-1) \tag{5·4}$$

and

$$X = \frac{(n-3)^2}{(n-2)} - t_{n-1}^2. \tag{5·5}$$

Having regard only to principal terms we find

$$1 - x^2 \simeq^{*} \frac{n(n-2)}{(n-1)^2}, \tag{5·6}$$

$$X \simeq 2n^{-2}(n-1)^3(n-2)^{-2}(n-3)(1-x''^2)\, v^{\frac{1}{2}}, \tag{5·7}$$

with

$$x' = \left(\frac{2}{3}\frac{n-2}{\overline{n-1}^2}\right)^{\frac{1}{2}} v^{\frac{1}{2}} x''. \tag{5·8}$$

Now, from (2·14),

$$f_n(t_n) = \left(\frac{n-1}{\pi n}\right)^{\frac{1}{2}} \frac{\frac{1}{2}(n-3)!}{\frac{1}{2}(n-4)!} \int_D dx(1-x^2)^{\frac{1}{2}(n-7)} f_{n-1}(t_{n-1}),$$

and, from the analysis in § 3, it will be clear that there are two separate parts of the domain D:

(I) a part near $x = 1/(n-1)$ for which t_{n-1} is entirely in the first zone and by hypothesis has the form (5·3);

(II) a part near $x = -1$ in which t_{n-1} assumes all values.

* The symbol $\simeq$ signifies 'equals, to required approximation'.

Let $$f_n(t_n) = f_n^{\mathrm{I}}(t_n) + f_n^{\mathrm{II}}(t_n), \tag{5·9}$$

where the functions on the right represent the contributions accruing from the respective parts of the domain of integration. Then

$$f_n^{\mathrm{I}}(t_n) \simeq \frac{\tfrac{1}{2}(n-3)!}{\tfrac{1}{2}(n-4)!\sqrt{\pi}}\left(\frac{n-1}{n}\right)^{\frac{1}{2}}\int_{x''=-1}^{x''=+1} dx(1-x^2)^{\frac{1}{2}(n-7)}C_{n-1}\{2n^{-2}(n-1)^3 \times (n-2)^{-2}(n-3)(1-x''^2)v^{\frac{1}{2}}\}^{\frac{1}{2}(n-5)},$$

which, on a change of variable from x to x'' by (5·4) and (5·8) and integrating in x'', becomes

$$\simeq C_{n-1}3^{-\frac{1}{2}}\frac{\tfrac{1}{2}(n-3)!\,\tfrac{1}{2}(n-5)!}{\tfrac{1}{2}(n-4)!^2}n^{-\frac{1}{2}(n-2)}(n-1)^{n-3}(n-2)^{-(n-4)}(n-3)^{\frac{1}{2}(n-5)}\left(\frac{\overline{n-2}^2}{n-1}-t_n^2\right)^{\frac{1}{2}(n-4)}, \tag{5·10}$$

which is of the required form. As regards the contribution of II, in (2·15) set

$$x+1 = \beta v + yv^{\frac{1}{2}}. \tag{5·11}$$

It will be found that when t_{n-1} has its limiting value $+(n-3)/(n-2)^{\frac{1}{2}}$ the vanishing of the terms in v^2 and v^3 gives

$$\beta = \tfrac{1}{3}n \quad \text{and} \quad y = \frac{2}{9}\sqrt{\frac{2}{3}\frac{(n-3)}{n^2(n-2)^{\frac{1}{2}}}},$$

whereas if t_{n-1} has its limiting value $-(n-3)/(n-2)^{\frac{1}{2}}$ the values are

$$\beta = \tfrac{1}{3}n \quad \text{and} \quad y = -\frac{2}{9}\sqrt{\frac{2}{3}\frac{(n-3)}{n^2(n-2)^{\frac{1}{2}}}}.$$

Between the limits of x,

$$-1+\frac{v}{3n}\pm\frac{2}{9}\sqrt{\frac{2}{3}\frac{n-3}{n^2(n-2)^{\frac{1}{2}}}}v^{\frac{1}{2}}, \tag{5·12}$$

$t_{n-1}(x)$ given by (2·15), assumes once all values between its limits of range and, in fact,

$$y \simeq \frac{2}{9}\sqrt{\frac{2}{3}\frac{t_{n-1}}{n^2}}. \tag{5·13}$$

Now $$f_n^{\mathrm{II}}(t_n) = \frac{\left(\frac{n-1}{n}\right)^{\frac{1}{2}}\tfrac{1}{2}(n-3)!}{\sqrt{\pi}\,\tfrac{1}{2}(n-4)!}\int dx(1-x^2)^{\frac{1}{2}(n-7)}f_{n-1}(t_{n-1}), \tag{5·14}$$

the limits of integration in x being given by (5·12). By (5·11) and (5·13) change the variable x into t_{n-1} (via y) when (5·14) becomes

$$f_n^{\mathrm{II}}(t_n) \simeq C(n)\,v^{\frac{1}{2}(n-4)}\int f_{n-1}(t_{n-1})\,dt_{n-1}, \tag{5·15}$$

and the integral on the right is unity. Written in full (5·15) then becomes

$$f_n^{\mathrm{II}}(t_n) \simeq \frac{1}{3\sqrt{\pi}\,n}\frac{\tfrac{1}{2}(n-3)!}{\tfrac{1}{2}(n-4)!}\frac{(n-1)^{\frac{1}{2}(n-3)}}{(3n\,\overline{n-2})^{\frac{1}{2}(n-4)}}\left(\frac{\overline{n-2}^2}{n-1}-t_n^2\right)^{\frac{1}{2}(n-4)}. \tag{5·16}$$

There is no difficulty now in proving by iteration from (5·10) and (5·16) that the constant has the form indicated in (5·1). Note that, in (5·9), f^{I} accounts for $(n-1)/n$ and f^{II} for $1/n$ of the total frequency.

6. Samples of 4

From (2·14), (2·15) and (3·7) the frequency for $n = 4$ is found to be

$$f_4(t_4) = \frac{\sqrt{3}}{2\pi}\int_D dx\, y^{-\frac{1}{2}}, \tag{6·1}$$

where $$y = 2(1-x^2)^3 - (v - 3x + 5x^3)^2 \quad \text{with} \quad v = \sqrt{3}\, t_4. \tag{6·2}$$

D is the range of values of x which give non-negative values for y with $|x| \leqslant 1$. Now

$$y = -(3x^2-1)^2(3x^2-2) - 2v(5x^3-3x) - v^2, \tag{6·3}$$

from which it appears that when v is small y has two real roots near $-1/\sqrt{3}$, two imaginary roots near* $+1/\sqrt{3}$, and single roots near $+\sqrt{2}/\sqrt{3}$ and $-\sqrt{2}/\sqrt{3}$ accounting thus for all six roots. With $O(v^{\frac{1}{2}}) = 0$ the four real roots are

$$-\frac{1}{\sqrt{3}} \pm \alpha v^{\frac{1}{2}} \quad \text{with} \quad \alpha^2 = \frac{2}{9\sqrt{3}}$$

$$\pm\sqrt{\frac{2}{3} - \frac{v}{9}}.$$

Hence the integral at (6·1) may be written as the sum of five integrals

$$\int_{-\sqrt{\frac{2}{3}}-v/9}^{-\sqrt{\frac{2}{3}}+v/9} + \int_{-\sqrt{\frac{2}{3}}+v/9}^{-1/\sqrt{3}-\alpha v^{\frac{1}{2}}} + \int_{-1/\sqrt{3}+\alpha v^{\frac{1}{2}}}^{1/\sqrt{3}-\alpha v^{\frac{1}{2}}} + \int_{1/\sqrt{3}-\alpha v^{\frac{1}{2}}}^{1/\sqrt{3}+\alpha v^{\frac{1}{2}}} + \int_{1/\sqrt{3}+\alpha v^{\frac{1}{2}}}^{\sqrt{\frac{2}{3}}-v/9}.$$

Fig. 2 illustrates the division of the region of integration D. There are five divisions, numbered I–V, in what follows, in which we regard as 'principal terms' only those in $\log v$ and the constant term. Terms in $v^{\frac{1}{2}}$ will ultimately be ignored:

$$\mathrm{I} = \int_{-\sqrt{\frac{2}{3}}-v/9}^{-\sqrt{\frac{2}{3}}+v/9} = O(v^{\frac{1}{2}}),$$

$$\mathrm{II} = \int_{-\sqrt{\frac{2}{3}}+v/9}^{-1/\sqrt{3}-\alpha v^{\frac{1}{2}}} \simeq \int dx(-1+3x^2)^{-1}(2-3x^2)^{-\frac{1}{2}}$$

$$\simeq \frac{1}{4\sqrt{3}}\left\{-4\left(\frac{\sqrt{6}}{9}v\right)^{\frac{1}{2}} - 3\alpha^2 v - \log 3\alpha^2 v\right\} \simeq \int_{1/\sqrt{3}+\alpha v^{\frac{1}{2}}}^{\sqrt{\frac{2}{3}}-v/9} \simeq \mathrm{V},$$

$$\mathrm{III} = \int_{-1/\sqrt{3}-\alpha v^{\frac{1}{2}}}^{+1/\sqrt{3}-\alpha v^{\frac{1}{2}}} \simeq \int dx(1-3x^2)^{-1}(2-3x^2)^{-\frac{1}{2}} \simeq -\frac{1}{2\sqrt{3}}\log\left(\frac{3\alpha^2 v}{1-3\alpha^2 v}\right),$$

$$\mathrm{IV} = \int_{1/\sqrt{3}-\alpha v^{\frac{1}{2}}}^{1/\sqrt{3}+\alpha v^{\frac{1}{2}}} \simeq 12^{-\frac{1}{2}}\int_{-1}^{+1}(x'^2+1)^{-\frac{1}{2}}\,dx' = \frac{1}{\sqrt{3}}\sinh^{-1}1.$$

Neglecting $O(v^{\frac{1}{2}})$ we accordingly have

$$\mathrm{I}+\mathrm{II}+\mathrm{III}+\mathrm{IV}+\mathrm{V} \simeq -\frac{1}{\sqrt{3}}\log 3\alpha^2 v + \frac{1}{\sqrt{3}}\sinh^{-1}1 + 2C. \tag{6·4}$$

The constant $2C$ derives from additional terms in integrals II, III and V:

$$\mathrm{III} = \int_{-\gamma}^{+\gamma} dx\{(1-3x^2)^2(2-3x^2) - \overbrace{2v(5x^3-3x)}^{z} - v^2\}^{-\frac{1}{2}}$$

with $$\gamma = 1/\sqrt{3} - \alpha v^{\frac{1}{2}}.$$

* In a sense which will be obvious from Fig. 2.

We have already taken account of the term in III found when v is zero. The constant C derives from the even powers of z in the formal expansion of the denominator of the integral element—the odd powers vanish by symmetry. Setting

$$x=\frac{1}{\sqrt{3}}-x',\quad 1-3x^2\simeq 2\sqrt{3}\,x',\quad 2-3x^2\simeq 1,\quad 5x^3-3x\simeq -\tfrac{4}{3}\sqrt{3}.$$

On expansion of III,

$$C\simeq 2\sum_{k=1}^{\infty}\int_{\alpha v^{\frac{1}{2}}}^{1/\sqrt{3}} dx'\,x'^{-4k-1}\,(2v)^{2k}\,(2\sqrt{3})^{-4k-1}\,C_{2k},$$

where C_{2k} are the even-order coefficients in the expansion of $(1+z)^{-\frac{1}{2}}$, i.e. C_{2k} is the coefficient of z^{2k}. On integration we are interested only in the value at the lower limit $\alpha v^{\frac{1}{2}}$, for all terms at the upper limit (and certain terms at the lower limit) are $O(v^{\frac{1}{2}})$ at least. Hence

$$C=\frac{1}{4\sqrt{3}}\sum_{k=1}^{\infty}C_{2k}/k.$$

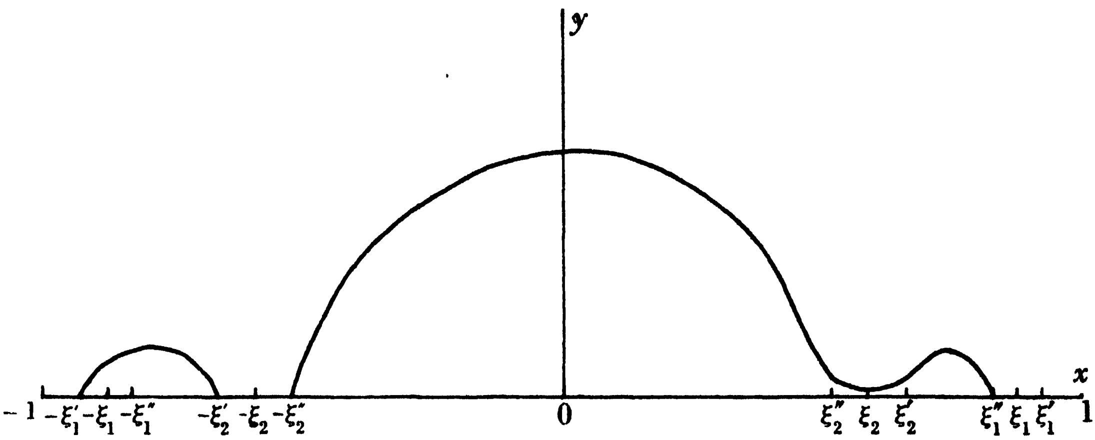


Fig. 2. Graph of y (see (6·2)).

Note. This diagram is designed merely to give a general idea of the limits of integration. It is not drawn to any scale. Following are the values of the ξ:

$$\begin{array}{ll} \xi_1=\sqrt{\tfrac{2}{3}} & \xi_2=1/\sqrt{3} \\ \xi_1'=\sqrt{\tfrac{2}{3}}+v/9 & \xi_2'=1/\sqrt{3}+\alpha v^{\frac{1}{2}} \\ \xi_1''=\sqrt{\tfrac{2}{3}}-v/9 & \xi_2''=1/\sqrt{3}-\alpha v^{\frac{1}{2}} \end{array}\Bigg\}\ \alpha^2=2/(9\sqrt{3})$$

Similarly II + V also yield C giving a constant additional term of $2C$.

Now
$$\begin{aligned} C&=\frac{1}{\sqrt{3}}\sum_{k=1}^{\infty}\frac{C_{2k}}{4k}=\frac{1}{\sqrt{3}}\int_0^1\frac{dx}{x}(C_2x^4+C_4x^8+\dots)\\ &=\frac{1}{\sqrt{3}}\int_0^1-\frac{dx}{x}+\frac{1}{2\sqrt{3}}\int_0^1\frac{dx}{x}\{(1+x^2)^{-\frac{1}{2}}+(1-x^2)^{-\frac{1}{2}}\}\\ &=\frac{1}{\sqrt{3}}\left[-\log x+\tfrac{1}{4}\log\frac{(1+x^2)^{\frac{1}{2}}-1}{(1+x^2)^{\frac{1}{2}}+1}+\tfrac{1}{4}\log\frac{1-(1-x^2)^{\frac{1}{2}}}{1+(1-x^2)^{\frac{1}{2}}}\right]_{x=0}^{x=1}\\ &=\frac{1}{\sqrt{3}}\{\log 2+\tfrac{1}{2}\log(2^{\frac{1}{2}}-1)\}. \end{aligned}\tag{6·5}$$

All logs are to base e (unless otherwise indicated in what follows). Hence

$$f_4(t_4) \simeq 0{\cdot}372646 - \frac{1}{2\pi}\log v \tag{6·6}$$

$$= 0{\cdot}285222 - 0{\cdot}366466\log_{10}|t_4|, \tag{6·7}$$

since $$v = \sqrt{3}\,t_4.$$

A. T. McKay (1933), from a different approach, gave the log term in (6·7), and, as a rough approximation to the constant term, the value 0·311568. He also showed that an expression of the form (6·7) accounted for most of the frequency, a fact of great importance. Assume that the residual term is of form

$$A|t_4|^{\frac{1}{2}} + B|t_4|,$$

and find A and B from

(i) $f_4(2/\sqrt{3}) = \frac{1}{2}\sqrt{3}$ (from (5·1); also McKay (1933)),

(ii) total frequency is unity,

giving

$$f_4(t_4) = 0{\cdot}285222 - 0{\cdot}366466\log_{10}|t_4| - 0{\cdot}009178|t_4|^{\frac{1}{2}} + 0{\cdot}031359|t_4|. \tag{6·8}$$

For algebraic manipulation at the next stage the form of residual $A'|t_4| + B't_4^2$ will be found more convenient, however, with A' and B' also determined from (i) and (ii). In this form

$$f_4(t_4) = 0{\cdot}285222 - 0{\cdot}159155\log|t_4| + 0{\cdot}014275|t_4| + 0{\cdot}007398t_4^2. \tag{6·9}$$

Note the smallness of the coefficients A, B, A' and B' in (6·8) and (6·9).

In the following table the first four even moments as derived from frequencies (6·8) and (6·9) are compared with the actual values as derived from the formulae (1·2). Both formulae yield excellent approximations, with (6·8) always superior to (6·9) however. Either formula can obviously be used with complete confidence for deriving the probability points. The frequency graph in Fig. 4 is derived from (6·8) which should also be used for the computation of the probability points.

Moment	Actual	Formula	
		(6·8)	(6·9)
μ_2	0·342857	0·342930	0·342470
μ_4	0·258941	0·258979	0·258606
μ_6	0·240503	0·240263	0·240205
μ_8	0·245940	0·245949	0·246662

7. Samples of 5

After many computational experiments the method used for determining the frequency $f_5(t_5)$ was as follows:

(1) Using (2·14) with form (6·9) for $f_4(t_4)$, central and link ordinates, i.e. $f_5(0)$ and $f_5(1/\sqrt{6})$ were computed.

(2) The approximate value of $f_5(t_5)$ near $t_5 = 1/\sqrt{6} + 0$ was found in the form

$$f_5(1/\sqrt{6}) + M(t_5 - 1/\sqrt{6})^{\frac{1}{2}},$$

M being known.

(3) The two zonal curves were found (i) passing through $(0, f_5(0))$ and $(1/\sqrt{6}, f_5(1/\sqrt{6}))$ with $f_5'(0) = 0$ and (ii) passing through $(1/\sqrt{6}, f_5(1/\sqrt{6}))$ and with the required form at $1/\sqrt{6}+0$ (i.e. as at (2) above) and at the limit $(\frac{3}{2}-0)$ so that $\mu_0\ (=1)$, μ_2, μ_4 and μ_6 have the exact values as given for $n = 5$ by the formulae (1·2).

Setting then

$$f_4(t_4) = 0{\cdot}285222 - 0{\cdot}159155 \log_e |t_4| + R(t_4), \tag{7·1}$$

with

$$t_4 = \frac{6}{\sqrt{5}} \frac{(x^3 - \frac{1}{2}x + \frac{1}{3}t_5)}{(1-x^2)^{\frac{3}{2}}}, \tag{7·2}$$

and

$$R(t_4) = 0{\cdot}014275\,|t_4| + 0{\cdot}007398 t_4^2, \tag{7·3}$$

we have

$$f_5(t_5) = \frac{4}{\pi\sqrt{5}} \int_\lambda^\mu \frac{dx}{1-x^2} f_4(t_4), \tag{7·4}$$

the limits of integration being λ (negative) and μ (positive) which are the values of x, from (7·2), corresponding respectively to $t_4 = -\frac{1}{2}\sqrt{3}$ and $t_4 = +\frac{1}{2}\sqrt{3}$. We shall be concerned only with the case $t_5 \leqslant 1/\sqrt{6}$ when $t_4(x)$ has three real roots β, α and γ of which β is negative and α and γ are positive. For (7·4) the following are required

$$\int_\lambda^\mu \frac{dx}{1-x^2} = \tfrac{1}{2}\log\left(\frac{1+\mu}{1+\lambda}\frac{1-\lambda}{1-\mu}\right), \tag{7·5}$$

$$\begin{aligned}\int_\lambda^\mu \frac{dx \log(1-x^2)}{(1-x^2)} &= \tfrac{1}{4}\{\log^2(1+\mu) - \log^2(1+\lambda) - \log^2(1-\mu) + \log^2(1-\lambda)\} \\ &\quad + \tfrac{1}{2}\{\log(1+\mu)\log(1-\mu) - \log(1+\lambda)\log(1-\lambda)\} \\ &\quad + \log 2 \log\left(\frac{1-\lambda}{1-\mu}\right) + J\left(\frac{1+\mu}{2}\right) - J\left(\frac{1+\lambda}{2}\right),\end{aligned} \tag{7·6}$$

$$\begin{aligned}\int_\lambda^\mu \frac{dx}{1-x^2}\log\left|x^3 - \frac{x}{2} + \frac{t_5}{3}\right| &= \tfrac{1}{2}\log\left(\frac{1-\lambda}{1-\mu}\right)\log(1-\alpha)(1-\beta)(1-\gamma) \\ &\quad - \tfrac{1}{2}\log\left(\frac{1+\lambda}{1+\mu}\right)\log(1+\alpha)(1+\beta)(1+\gamma) + \frac{1}{2}\left\{\sum_{i=1}^{6} I(\kappa_i) + \sum_{j=7}^{12} J(\kappa_j)\right\},\end{aligned} \tag{7·7}$$

with

$$\begin{aligned}
\kappa_1 &= (\alpha-\lambda)/(1-\alpha), & \kappa_5 &= (\beta-\lambda)/(1-\beta), & \kappa_9 &= (\gamma-\lambda)/(1+\gamma),\\
\kappa_2 &= (\mu-\alpha)/(1+\alpha), & \kappa_6 &= (\mu-\beta)/(1+\beta), & \kappa_{10} &= (\mu-\gamma)/(1-\gamma),\\
\kappa_3 &= (\gamma-\lambda)/(1-\gamma), & \kappa_7 &= (\alpha-\lambda)/(1+\alpha), & \kappa_{11} &= (\beta-\lambda)/(1+\beta),\\
\kappa_4 &= (\mu-\gamma)/(1+\gamma), & \kappa_8 &= (\mu-\alpha)/(1-\alpha), & \kappa_{12} &= (\mu-\beta)/(1-\beta),
\end{aligned}$$

and

$$I(\kappa) = \int_0^\kappa \frac{dx \log x}{1+x} = \log\kappa \log(1+\kappa) - \psi(\kappa),$$

$$J(\kappa) = \int_0^\kappa \frac{dx \log x}{1-x} = -\log\kappa\log(1-\kappa) - \phi(\kappa),$$

$$\left.\begin{aligned}\phi(\kappa) &= \frac{\kappa}{1^2} + \frac{\kappa^2}{2^2} + \frac{\kappa^3}{3^2} + \dots \\ \psi(\kappa) &= \frac{\kappa}{1^2} - \frac{\kappa^2}{2^2} + \frac{\kappa^3}{3^2} - \dots\end{aligned}\right\} \text{ when } \kappa \leqslant 1.$$

It is useful to note that $\phi(1) = 1{\cdot}644934 = 2\psi(1)$. The functions $\phi(\kappa)$ and $\psi(\kappa)$ do not appear to be tabulated. By fitting curves to their values for equally spaced intervals of 0·05 from 0 to 0·5 the following very close approximations are found, applicable for $x \leqslant \frac{1}{2}$:

$$\phi(x) = 1{\cdot}000567x + 0{\cdot}233454x^2 + 0{\cdot}186052x^3,$$
$$\psi(x) = 0{\cdot}999835x - 0{\cdot}244220x^2 + 0{\cdot}077024x^3.$$

When $1 \geqslant \kappa > \frac{1}{2}$ the following formulae can be used:

$$\phi(\kappa) = \phi(1) - \log\kappa \log(1-\kappa) - \phi(1-\kappa),$$
$$\psi(\kappa) = \tfrac{1}{2}\phi(1) + \log\kappa\log(1+\kappa) - \phi(1-\kappa) + \tfrac{1}{2}(1-\kappa^2).$$

When $\kappa > 1$ we use

$$\phi(\kappa) = 2\phi(1) - \tfrac{1}{2}\log^2 1/\kappa - \phi(1/\kappa),$$
$$\psi(\kappa) = 2\psi(1) + \tfrac{1}{2}\log^2 1/\kappa - \psi(1/\kappa).$$

Another useful formula is

$$\phi(\kappa) = \psi(\kappa) + \tfrac{1}{2}\phi(\kappa^2).$$

The algebra of the contribution to (7·4) from $R(t_4)$ is without mathematical interest. From the formula the following were the values found for the central frequency and the second link frequency:

$$f_5(0) = 0{\cdot}606563; \quad f_5(1/\sqrt{6}) = 0{\cdot}599069. \tag{7·8}$$

The moments, computed by (2·1), with $n = 5$, are

$$\mu_0 = 1, \quad \mu_2 = 0{\cdot}375, \quad \mu_4 = 0{\cdot}361607, \quad \mu_6 = 0{\cdot}474609, \quad \mu_8 = 0{\cdot}719382. \tag{7·9}$$

Computation by approximate integration of certain of the ordinates gave evidence of marked irregularity near the link $t_5 = 1/\sqrt{6}$. In consequence, it seemed desirable to try to find a term (in addition to the constant given at (7·8)) of the expansion of $f_5(t_5)$ near $1/\sqrt{6}+0$. Setting

$$t_5 = \frac{1}{\sqrt{6}} + t, \tag{7·10}$$

where t is small and positive—we shall be interested only in a term in $t^{\frac{1}{2}}$—we find

$$f_5(t_5) \simeq \frac{4}{\pi\sqrt{5}}\left\{\int_{\lambda+\lambda' t}^{v-At^{\frac{1}{2}}} + \int_{v+At^{\frac{1}{2}}}^{\mu+\mu' t}\right\}\frac{dx}{1-x^2} f_4(t_4). \tag{7·11}$$

The values $v \pm At^{\frac{1}{2}}$ are the abscissae of the points at which the curve $t_4 = t_4(x)$, given by (7·2), intersects the t_4 link line $t_4 = 2/\sqrt{3}$ near $x = v = -\frac{1}{4}\sqrt{6}$. It can easily be shown that

$$A^2 = \frac{\sqrt{6}}{12}. \tag{7·12}$$

We are not concerned with the values $(\lambda + \lambda' t)$ and $(\mu + \mu' t)$ which are the abscissae corresponding to the intersection of $t_4 = t_4(x)$ with $t_4 = -2/\sqrt{3}$ and its third intersection with $t_4 = 2/\sqrt{3}$. Remembering that at the latter link $f_4(t_4)$ has the value $1/(2\sqrt{3})$, the *integral-free* term in $t^{-\frac{1}{2}}$ in the first derivative $f_5'(t_5)$ of $f_5(t_5)$ is

$$\frac{4}{\pi\sqrt{5}}\frac{1}{1-v^2}\{f_4(v-At^{\frac{1}{2}}) \times -\tfrac{1}{2}At^{-\frac{1}{2}} + f_4(v+At^{\frac{1}{2}}) \times -\tfrac{1}{2}At^{-\frac{1}{2}}\} = -\frac{16A\sqrt{15}}{75\pi}t^{-\frac{1}{2}}. \tag{7·13}$$

Also we have to consider the *integral* term in $f_5'(t_5)$. For this purpose, from (7·10) and (7·2),

$$t_4 = \frac{6}{\sqrt{5}}\left\{(x+\sqrt{\tfrac{2}{3}})\left(x-\frac{1}{\sqrt{6}}\right)^2 + \frac{t}{3}\right\}(1-x^2)^{-\frac{3}{2}}$$
$$\simeq \frac{6}{\sqrt{5}}\left(x+\sqrt{\tfrac{2}{3}}+\frac{2t}{9}\right)\left\{\left(x-\frac{1}{\sqrt{6}}-\frac{t}{9}\right)^2 + \frac{\sqrt{6}}{9}t\right\}(1-x^2)^{-\frac{3}{2}}.$$

Remembering (7·1), it can be shown that the only term in (7·11) from which a term in $t^{\frac{1}{2}}$ can come is approximately

$$-\frac{4}{\pi\sqrt{5}}\frac{1}{2\pi}\int_{\lambda}^{\mu}\frac{dx}{1-x^2}\log\left\{\left(x-\frac{1}{\sqrt{6}}-\frac{t}{9}\right)^2+\frac{\sqrt{6}}{9}t\right\}, \tag{7·14}$$

λ and μ, the limiting values, being respectively negative and positive.

Differentiating (7·14) in respect of t we find

$$-\frac{2}{\pi^2\sqrt{5}}\int_{\lambda}^{\mu}\frac{dx}{1-x^2}\left\{\left(x-\frac{1}{\sqrt{6}}-\frac{t}{9}\right)^2+\frac{\sqrt{6}}{9}t\right\}^{-1}\left\{\frac{\sqrt{6}}{9}-\frac{2}{9}\left(x-\frac{1}{\sqrt{6}}-\frac{t}{9}\right)\right\}. \tag{7·15}$$

Changing variables by $$x-\left(\frac{1}{\sqrt{6}}+\frac{t}{9}\right)=\frac{6^{\frac{1}{4}}}{3}t^{\frac{1}{2}}y$$

and letting t tend towards $+0$, we find for the term in $t^{-\frac{1}{2}}$

$$-\frac{2}{\pi^2\sqrt{5}}\left(\frac{\pi 6^{\frac{1}{4}}2}{5}t^{-\frac{1}{2}}\right). \tag{7·16}$$

Adding (7·13) and (7·16) we find

$$-\frac{1}{\pi}5^{-\frac{1}{2}}2^{\frac{9}{4}}3^{-\frac{3}{4}}t^{-\frac{1}{2}}.$$

On integrating we find for the term in $t^{\frac{1}{2}}=\left(t_5-\frac{1}{\sqrt{6}}\right)^{\frac{1}{2}}$

$$-\frac{1}{\pi}5^{-\frac{1}{2}}2^{\frac{13}{4}}3^{-\frac{3}{4}}\left(t_5-\frac{1}{\sqrt{6}}\right)^{\frac{1}{2}}=-0·594117\left(t_5-\frac{1}{\sqrt{6}}\right)^{\frac{1}{2}}. \tag{7·17}$$

From (5·1) the value of $f_5(x)$ near $x=\frac{3}{2}$ is

$$0·219166(\tfrac{3}{2}-x)^{\frac{1}{2}}, \tag{7·18}$$

where x is usually written for simplicity instead of t_5 in the remainder of this section.

Having regard to (7·8) and to the fact that, from § 4, $f_5'(0)=0$, in the half-zone $(0-1/\sqrt{6})$, $f_5(x)=F(x)$ must be of form

$$F(x)=0·606563+a_2x^2+a_3x^3+a_4x^4. \tag{7·19}$$

The first relation between the coefficients is found by giving expression to the fact that $y=F(x)$ passes through the link-point $(1/\sqrt{6}, 0·599069)$:

$$0·166667a_2+0·068041a_3+0·027778a_4=-0·007494. \tag{7·20}$$

In the zone $(1/\sqrt{6}-\frac{3}{2})$ assume that

$$f_5(x)=G(x)=-0·594117\left(x-\frac{1}{\sqrt{6}}\right)^{\frac{1}{2}}+0·219166(\tfrac{3}{2}-x)^{\frac{1}{2}}+b_0+b_1(\tfrac{3}{2}-x)$$
$$+b_2(\tfrac{3}{2}-x)^2+b_3(\tfrac{3}{2}-x)^3, \tag{7·21}$$

designed to conform with requirements (7·17) and (7·18). Since $y=G(x)$ must pass through $(\frac{3}{2}, 0)$,

$$b_0=0·594117\left(\tfrac{3}{2}-\frac{1}{\sqrt{6}}\right)^{\frac{1}{2}}=0·620775. \tag{7·22}$$

Taking the value of b_0 into account and giving algebraic expression to $y=G(x)$ passing through $(1/\sqrt{6}, 0·599069)$, we find

$$1·091752b_1+1·191922b_2+1·301283b_3=-0·250706. \tag{7·23}$$

To find the six coefficients a_2, a_3, a_4 (in (7·19)) and b_1, b_2, b_3 (in (7·21)), we have, so far, found two equations, (7·20) and (7·23). The remaining four equations are found by equating the total frequency to unity and the first three even moments to their true values given at (7·9), i.e. setting

$$\tfrac{1}{2}\mu_{2k} = \int_0^{1/\sqrt{6}} dx\, x^{2k} F(x) + \int_{1/\sqrt{6}}^{\frac{3}{2}} dx\, x^{2k}\, G(x) \quad (k = 0, 1, 2, 3).$$

On substituting for a_3 given by (7·20), for b_0 given by (7·22) and for b_3 given by (7·23), we find the four equations in a_2, a_3, b_1, b_2:

$$\left.\begin{aligned}
0\cdot0^{2}90721a_2 + 0\cdot0^{2}139999a_3 + 0\cdot297981b_1 + 0\cdot108441b_2 &= -0\cdot071173,\\
0\cdot0^{3}64802a_2 + 0\cdot0^{3}110237a_3 + 0\cdot268332b_1 + 0\cdot082590b_2 &= -0\cdot066293,\\
0\cdot0^{4}5901a_2 + 0\cdot0^{4}107175a_3 + 0\cdot303248b_1 + 0\cdot079086b_2 &= -0\cdot076019,\\
0\cdot0^{5}6368a_2 + 0\cdot0^{5}1169a_3 + 0\cdot400090b_1 + 0\cdot089674b_2 &= -0\cdot101300.
\end{aligned}\right\} \qquad (7\cdot24)$$

On solution (and checking by substitution) the coefficients are found to give finally the following frequencies:

$$\left.\begin{aligned}
&\text{Zone} && f_5(x)\\
&0 - 1/\sqrt{6}: && 0\cdot606563 - 0\cdot3307x^2 + 3\cdot1955x^3 - 6\cdot1129x^4,\\
&1/\sqrt{6} - \tfrac{3}{2}: && 0\cdot620775 - 0\cdot594117\left(x - \frac{1}{\sqrt{6}}\right)^{\frac{1}{2}} + 0\cdot219166(\tfrac{3}{2} - x)^{\frac{1}{2}} - 0\cdot268273(\tfrac{3}{2} - x)\\
& && + 0\cdot067263(\tfrac{3}{2} - x)^2 - 0\cdot029195(\tfrac{3}{2} - x)^3,
\end{aligned}\right\} \qquad (7\cdot25)$$

with $x = t_5$.

The extremely interesting form of the frequency curve may be observed from Fig. 5. In the first half-zone the frequency shows but little variation: the curve declines to a minimum of 0·6058 at $x = 0\cdot0894$ then rises to a maximum of 0·6136 at $x = 0\cdot3027$. It then recedes to the link $1/\sqrt{6}$, where it assumes the value 0·5991. As one type of check on the reliability of the results in general, some ordinates were computed directly (i.e. using (7·4) and (7·1)), or by approximate integration using (7·4) and (6·8) and compared with the ordinates computed from (7·25) to the following effect:

Trial value of t_5	Value of frequency	
	By approx. integration	By (7·25)
0·15	0·6069	0·6068
0·3	0·6106	0·6136
0·6	0·3650	0·3603
0·9	0·2232	0·2308
1·2	0·1377	0·1371

Except perhaps for the frequency at $t_5 = 0\cdot9$, the correspondence is satisfactory; there can be little doubt that the more accurate figures are those from (7·25).

As a stringent test of the accuracy of the frequency the 8th moment μ_8' was computed from the empirical curves at (7·25) and compared with the actual value given at (7·9):

$$\mu_8' = 0\cdot7191, \quad \mu_8 = 0\cdot7194,$$

Even as the figures stand the check is decisive: it should be added that the 4th place of decimals in μ_8' is suspect to the approximation used.

8. SAMPLES OF 6

In this case the links are 0, $1/\sqrt{2}$ and $4/\sqrt{5}$, and the link frequencies at the first two were found by approximate integration using form (2·13) with t_6 given by (2·8). For this purpose, drawings were made of the two sections of $f_5(t_5)$ on a scale sufficient to ensure that an ordinate read for any abscissa would be correct probably to the 3rd place of decimals. For intervals of 1°, values of t_5 were computed over the whole range by (2·8) (for t_6 given), and graphically the value of $f_5(t_5)$ was read off for each t_5. Hundreds of readings had to be made, but actually the work, with a little practice, was rapid and accurate, the entries being practically self-checking. The Gregory formula (using 2 correction terms) was used to give the following results:

$$f_6(0) = 0{\cdot}6889; \quad f_6(1/\sqrt{2}) = 0{\cdot}3247. \tag{8·1}$$

The two zonal frequency curves, say $y = F(x)$ in $(0-1/\sqrt{2})$ and $y = G(x)$ in $(1/\sqrt{2}-4/\sqrt{5})$, *writing x instead of t_6* must have the following properties:

$$\left.\begin{array}{ll}
\text{(i)} & F(0) = 0{\cdot}6889,\\
\text{(ii)} & F'(0) = 0 \ (\S\,4),\\
\text{(iii)} & F(1/\sqrt{2}) = G(1/\sqrt{2}) = 0{\cdot}3247,\\
\text{(iv)} & F'(1/\sqrt{2}) = G'(1/\sqrt{2}) \ (\S\,3),\\
\text{(v)} & G(x) \simeq 5(4/\sqrt{5}-x)/36 \text{ (from (5·1))}.
\end{array}\right\} \tag{8·2}$$

The curves were

$$\left.\begin{aligned}
F(x) &= 0{\cdot}6889 + a_2x^2 + a_3x^3 + a_4x^4 + a_5x^5,\\
G(x) &= \tfrac{5}{36}(\beta-x) + b_2(\beta-x)^2 + b_3(\beta-x)^3 + b_4(\beta-x)^4 \text{ with } \beta = 4/\sqrt{5}.
\end{aligned}\right\} \tag{8·3}$$

The exact moments are

$$\mu_0 = 1, \quad \mu_2 = 0{\cdot}380952, \quad \mu_4 = 0{\cdot}409191, \quad \mu_6 = 0{\cdot}642924, \quad \mu_8 = 1{\cdot}219892. \tag{8·4}$$

It is proposed to compute the seven coefficients in (8·3) using (8·2) and (8·4). Now, with a curve of the type of $f_6(x)$, where much of the frequency is at the ends it is evident that the contribution from the zone $(0-1/\sqrt{2})$ to the higher moments μ_6 and μ_8 is exceedingly minute: this property was utilized to divide the single series of seven equations into two series of three and four equations using the following device: Approximate $F(x)$ by a curve $F_1(x)$ given by

$$F_1(x) = 0{\cdot}6889 + a_2'x^2, \tag{8·5}$$

finding a_2' simply by passing $y = F_1(x)$ through $(1/\sqrt{2}, 0{\cdot}3247)$ giving $a_2' = -0{\cdot}7284$. If μ_{2s} is the moment, let μ_{2s}' and μ_{2s}'' be the contributions from $F(x)$ and $G(x)$ respectively so that $\mu_{2s} = \mu_{2s}' + \mu_{2s}''$. Let ν_{2s}' be the estimate, using $F_1(x)$, of μ_{2s}'. For $s = 2, 3, 4$ the values are

$$\nu_4' = 0{\cdot}030318, \quad \nu_6' = 0{\cdot}009360, \quad \nu_8' = 0{\cdot}003397,$$

which, subtracted from the corresponding μ_{2s} given by (8·4), give very close estimates of μ_{2s}'', which involve only b_2, b_3, b_4. The equations in order

$$\left.\begin{array}{ll}
(1) & 1{\cdot}170178b_2 + 1{\cdot}265837b_3 + 1{\cdot}369316b_4 = 0{\cdot}174500,\\
(2) & 0{\cdot}726980b_2 + 0{\cdot}378949b_3 + 0{\cdot}235502b_4 = 0{\cdot}058771,\\
(3) & 1{\cdot}260067b_2 + 0{\cdot}534702b_3 + 0{\cdot}289663b_4 = 0{\cdot}091217,
\end{array}\right\} \tag{8·6}$$

are found from (iii) at (8·2), from μ_6'' and from μ_8''.

The equations in the a's are

$$\left.\begin{aligned}
&(4)\quad 0{\cdot}5a_2 \qquad\qquad + 0{\cdot}353553a_3 + 0{\cdot}25a_4 \qquad + 0{\cdot}176777a_5 = -0{\cdot}3642,\\
&(5)\quad 1{\cdot}414214a_2 + 1{\cdot}5a_3 \qquad + 1{\cdot}414214a_4 + 1{\cdot}25a_5 \qquad = -0{\cdot}653179,\\
&(6)\quad 0{\cdot}117851a_2 + 0{\cdot}0625a_3 \quad + 0{\cdot}035355a_4 + 0{\cdot}020833a_5 = -0{\cdot}115790,\\
&(7)\quad 0{\cdot}035355a_2 + 0{\cdot}020833a_3 + 0{\cdot}012627a_4 + 0{\cdot}007813a_5 = -0{\cdot}031433,
\end{aligned}\right\} \qquad (8{\cdot}7)$$

where (4) is from (iii) at (8·2), (5) from (iv) at (8·2), (6) from the total frequency $= \frac{1}{2}$ and (7) from variance $= \mu_2$.

The solutions of (8·6) and (8·7) yield the following frequencies:

$$\left.\begin{aligned}
&\text{Zone} && f_6(x)\\
&0 - 1/\sqrt{2} && : 0{\cdot}6889 - 1{\cdot}2715x^2 - 2{\cdot}6073x^3 + 9{\cdot}5669x^4 - 6{\cdot}7790x^5,\\
&1/\sqrt{2} - 4/\sqrt{5} && : \tfrac{5}{36}(\beta - x) + 0{\cdot}047068(\beta - x)^2 + 0{\cdot}024897(\beta - x)^3 + 0{\cdot}064198(\beta - x)^4,\\
& && \qquad\qquad\qquad\qquad \text{with } \beta = 4/\sqrt{5},
\end{aligned}\right\} \qquad (8{\cdot}8)$$

with $x = t_6$.

As a check, the 4th moment computed from the foregoing curves gave 0·4108 as compared with the actual $\mu_4 = 0{\cdot}4092$, an error of 0·38 %. This is not of any importance from the viewpoint of the computation of the probability points, but it illustrates how, using the integral iteration method as generally in this paper, the momental check reveals increasing discrepancies with increasing n.

It might be thought that by constructing empirically 'almost any' symmetrical frequency curve, so that say the 0th, 2nd and 4th moments have the true values, we shall ensure that the subsequent even moments computed from such an empirical curve will approximate closely to the corresponding true values. That this is not the case may be seen by computing the 6th and 8th moments by the well-known Karl Pearson iteration formula,* where μ_2 and μ_4 have their true values, for $\sqrt{b_1}$ with $n = 6$:

	Actual	Karl Pearson iteration	Percentage discrepancy
$\beta_4 = \mu_6/\mu_2^3$	11·6291	12·4984	+10·7
$\beta_6 = \mu_8/\mu_2^4$	57·9214	73·3990	+26·7

Even when, in the Pearson iteration for β_6, one gives β_4 the correct value, we find a percentage discrepancy of 17·9. These percentages place in perspective the minuteness of the percentage errors found in using the higher momental check as it is used throughout this paper.

It is an interesting question of general import whether in work of this kind the arduous and potentially erroneous computation (by integral iteration) of the central and link frequencies could be dispensed with, and reliance placed entirely on the moments, together with the functional properties of the frequencies, which, of course, merely represent an elaboration of the Karl Pearson approach. In this connexion a couple of experiments were made on the $\sqrt{b_1}$ frequency for $n = 6$.

* *Tables for Statisticians and Biometricians*, Part I, 2nd ed., p. xi.

For the first experiment, the two zonal curves were assumed to have the correct order of contact, the correct form, (8·2) (v), at the limit of range and the correct values of μ_0 ($=1$), μ_2, μ_4 and μ_6. The equations are

$$\left.\begin{aligned}
&\text{Zone}\\
&0-1/\sqrt{2} \quad : F_1(x) = 0\cdot659844-1\cdot075618x^2+0\cdot555991x^3,\\
&1/\sqrt{2}-4/\sqrt{5} : G_1(x) = \tfrac{5}{36}(\beta-x)+0\cdot080560(\beta-x)^2-0\cdot085469(\beta-x)^3\\
&\qquad\qquad +0\cdot133119(\beta-x)^4 \text{ with } \beta = 4/\sqrt{5}.
\end{aligned}\right\} \tag{8·9}$$

This gives a central frequency 0·6598 compared with the computed frequency (by (8·1)) of 0·6889. In all the circumstances the difference is not important. The 8th moment, μ_8'', from (8·9), is 1·217706, or $-0\cdot18$ % in error.

The second experiment contemplated the frequency as a single-curve system with correct first derivative ($-\frac{5}{36}$) at the limit and with correct μ_0, μ_2, μ_4, μ_6. The curve is

$$F_2(x) = 0\cdot669426-1\cdot51097x^2+1\cdot53854x^3-0\cdot60545x^4+0\cdot085157x^5, \tag{8·10}$$

which has the properties: (i) the central ordinate 0·6694 is close to the actual; (ii) limit value from curve scarcely differed from the actual since $F_2(4/\sqrt{5}) = 0\cdot0015$; (iii) μ_8''' from curve $= 1\cdot2237$, an error of 0·31 %.

All the systems (8·8), (8·9) or (8·10) yield probability points which differ very little. For instance, in the three cases, the 5 % point is given by

System	5 % probability
(8·8)	1·0432
(8·9)	1·0385
(8·10)	1·0384

The practical identity of the latter two is due to the fact that the frequencies were derived on very similar hypotheses: it does not mean that the result is more reliable than that from (8·8) which, assuming the accuracy of the calculation of the link ordinates, must be deemed to be the most correct and is adopted for the iteration to the $n = 7$ stage. Nevertheless, these experiments convey the hint of general application that if we know (i) a number of moments, (ii) the limits of range and the frequency form at the limits of range, and (iii) that the amount of frequency near the limits of range is not negligible, we will probably be in a position to estimate with fair accuracy the points of low probability. For this, however, hypothesis (iii) is essential: it has no value from the computational point of view if the frequency near the limits is negligible. This point is discussed further in § 10.

9. Samples of 7

The functional properties of the curves at the stage are as follows. Let the three links be denoted by α, β, γ, so that

$$\alpha = 1/\sqrt{12}, \quad \beta = 3/\sqrt{10}, \quad \gamma = 5/\sqrt{6}. \tag{9·1}$$

Denoting t_7 by x, set

$$y = x-\alpha, \quad z = \gamma-x, \tag{9·2}$$

and let the curves in the half-zone $(0-1/\sqrt{12})$, and in the zones $(1/\sqrt{12}-3/\sqrt{10})$ and $(3/\sqrt{10}-5/\sqrt{6})$ be denoted respectively by $F(x)$, $G(y)$ and $H(z)$. We then have

$$\left.\begin{aligned}
&\text{(i)}\ F(0) = 0{\cdot}6781 = A,\\
&\text{(ii)}\ F'(0) = 0,\\
&\text{(iii)}\ F(\alpha) = G(0) = 0{\cdot}5870 = B,\\
&\text{(iv)}\ F'(\alpha) = G'(0),\\
&\text{(v)}\ F''(\alpha) = G''(0),\\
&\text{(vi)}\ G(\beta-\alpha) = H(\gamma-\beta) = 0{\cdot}1838 = C,\\
&\text{(vii)}\ G'(\beta-\alpha) = -H'(\gamma-\beta),\\
&\text{(viii)}\ H(z) = Dz^{\frac{5}{2}}+c_2z^2+c_3z^3 \text{ with } D = 0{\cdot}078091.
\end{aligned}\right\} \qquad (9{\cdot}3)$$

The central and link ordinates A, B and C at (i), (iii) and (vi), were derived by the Gregory formula from (2·14), using intervals of 0·01, 0·025 and 0·05 at different sections of the integral range. The equalities in the derivatives at the links are in accordance with order of contact requirements (§ 3). The first term on the right of (viii) is from (5·1) with $n = 7$.

Conditions (9·3) determine the form of the polynomials:

$$\left.\begin{aligned}
F(x) &= A+a_2x^2+a_4x^4,\\
G(y) &= B+(2a_2\alpha+4a_4\alpha^3)\,y+\tfrac{1}{2}(2a_2+12a_4\alpha^2)\,y^2+b_3y^3+b_4y^4,\\
H(z) &= Dz^{\frac{5}{2}}+c_2z^2+c_3z^3,
\end{aligned}\right\} \qquad (9{\cdot}4)$$

with $x = t_7$.

$F(x)$ is taken as an even function of x because it is symmetrical in the zone $(-1/\sqrt{12}$ to $+1/\sqrt{12})$. This should have been done in the case of $n = 5$; neglect to do so was not serious enough to render recalculation necessary.

The moments used were:

$$\mu_0 = 1, \quad \mu_2 = 0{\cdot}375, \quad \mu_4 = 0{\cdot}421875, \quad \mu_6 = 0{\cdot}733487. \qquad (9{\cdot}5)$$

Using (9·3) in conjunction with μ_0 and μ_2 (only) in (9·5) the following equations in the six unknowns a_2, a_4, b_3, b_4, c_2, c_3 were found:

Eqn. no.	Left: coefficients of						Right: absolute term
	a_2	a_4	b_3	b_4	c_2	c_3	
1	12	1	—	—	—	—	−13·112496
2	0·816667	0·281315	0·287507	0·189756	—	—	−0·403210
3	—	—	—	—	1·193685	1·304172	0·094626
4	1·897366	0·756233	1·306833	1·150028	2·185118	3·581055	−0·122438
5	0·229605	0·069278	0·047439	0·025048	0·434724	0·356221	−0·122152
6	0·130637	0·041871	0·032192	0·017835	0·668438	0·496634	−0·044365

Approximations to F, G and H were found:

$$\left.\begin{aligned}
F_1(x) &= 0{\cdot}6781-1{\cdot}238888x^2+1{\cdot}754160x^4,\\
G_1(y) &= 0{\cdot}5870-0{\cdot}546478y-0{\cdot}361808y^2+0{\cdot}171129y^3+0{\cdot}347163y^4,\\
H_1(z) &= 0{\cdot}078091z^{\frac{5}{2}}-0{\cdot}017319z^2+0{\cdot}088408z^3.
\end{aligned}\right\} \qquad (9{\cdot}6)$$

These yielded estimates of the 4th and 6th moments as follows:

$$\mu_4' = 0{\cdot}419712, \quad \mu_6' = 0{\cdot}720776, \tag{9·7}$$

differing by $-0{\cdot}5\,\%$ and $-1{\cdot}7\,\%$ respectively from the correct values at (9·5). These deviations were not serious from the viewpoint of probability-point determination. Nevertheless, it seemed worth while to try to achieve a closer approximation. This was done by finding a 'corrector' $\phi(x)$ (not positive, like a frequency, for all values of x) with the following properties:

$$\left.\begin{array}{ll} \text{(i)} & \text{total 'frequency' zero,} \\ \text{(ii)} & \text{'2nd moment' zero,} \\ \text{(iii)} & \phi'(0) = 0, \\ \text{(iv)} & \phi(\gamma) = \phi'(\gamma) = 0, \\ \text{(v)} & \text{'4th moment'} = \mu_4 - \mu_4' = 0{\cdot}002163. \end{array}\right\} \tag{9·8}$$

Then $$\phi(x) = 0{\cdot}002404 - 0{\cdot}028853x^2 + 0{\cdot}045232x^3 - 0{\cdot}024236x^4 + 0{\cdot}004342x^5, \tag{9·9}$$

and the frequencies finally adopted are

$$\left.\begin{array}{lll} 0 - 1/\sqrt{12} & \ldots & F(x) = F_1(x) + \phi(x), \\ 1/\sqrt{12} - 3/\sqrt{10} & \ldots & G(y) = G_1(y) + \phi(x), \\ 3/\sqrt{10} - 5/\sqrt{6} & \ldots & H(z) = H_1(z) + \phi(x), \end{array}\right\} \tag{9·10}$$

F_1, G_1 and H_1 being given by (9·6) and $x = t_7$. It is evident from the smallness of the coefficients of $\phi(x)$ in (9·9) that the correction effected by $\phi(x)$ is minute. From (9·10) the moment μ_6'' is 0·728972, so that the error is reduced to about one-third of what it was using F_1, G_1 and H_1.

10. SAMPLES OF 8

The links and link frequencies are as follows:

$$\left.\begin{array}{rcl} \text{Link} & & \text{Link frequency} \\ 0 & : & 0{\cdot}6927 = A, \\ \beta = 2/\sqrt{15} & : & 0{\cdot}4442 = B, \\ \gamma = 2/\sqrt{3} & : & 0{\cdot}1018 = C, \\ \delta = 6/\sqrt{7} & : & 0{\cdot}04019153(\delta - x)^2 = D(\delta - x)^2/2, \end{array}\right\} \tag{10·1}$$

where $x = t_8$.

Set

$$\left.\begin{array}{l} y = x - \beta, \\ z = \delta - x, \\ \kappa = \gamma - \beta = 0{\cdot}638303, \\ \lambda = \delta - \gamma = 1{\cdot}113086. \end{array}\right\} \tag{10·2}$$

The orders of contact (§3) entail the following forms for the three zones:

$$\left.\begin{array}{ll} \text{Zone} & \\ 0 - 2/\sqrt{15} : & F(x) = A + a_2\dfrac{x^2}{2} + a_3\dfrac{x^3}{6} + a_4\dfrac{x^4}{24}, \\ 2/\sqrt{15} - 2/\sqrt{3} : & G(y) = B + \left(a_2\beta + a_3\dfrac{\beta^2}{2} + a_4\dfrac{\beta^3}{6}\right)y + \left(a_2 + a_3\beta + a_4\dfrac{\beta^2}{2}\right)\dfrac{y^2}{2} + b_3\dfrac{y^3}{6} + b_4\dfrac{y^4}{24}, \\ 2/\sqrt{3} - 6/\sqrt{7} : & H(z) = D\dfrac{z^2}{2} + c_3\dfrac{z^3}{6} + c_4\dfrac{z^4}{24}. \end{array}\right\} \tag{10·3}$$

Five of the seven equations required to determine the a, b and c will be found from the order of contact conditions, as follows:

$$\text{(i)}\quad B = A + a_2\frac{\beta^2}{2} + a_3\frac{\beta^3}{6} + a_4\frac{\beta^4}{24},$$

$$\text{(ii)}\quad C = B + \left(a_2\beta + a_3\frac{\beta^2}{2} + a_4\frac{\beta^3}{6}\right) + \left(a_2 + a_3\beta + a_4\frac{\beta^2}{2}\right)\frac{\kappa^2}{2} + b_3\frac{\kappa^3}{6} + b_4\frac{\kappa^4}{24},$$

$$\text{(iii)}\quad C = D\frac{\lambda^2}{2} + c_3\frac{\lambda^3}{6} + c_4\frac{\lambda^4}{24},$$

$$\text{(iv)}\quad a_2\beta + a_3\frac{\beta^2}{2} + a_4\frac{\beta^3}{6} + \left(a_2 + a_3\beta + a_4\frac{\beta^2}{2}\right)\kappa + b_3\frac{\kappa^2}{2} + b_4\frac{\kappa^3}{6} = -D\lambda - c_3\frac{\lambda^2}{2} - c_4\frac{\lambda^3}{6},$$

$$\text{(v)}\quad a_2 + a_3\beta + a_4\frac{\beta^2}{2} + b_3\kappa + b_4\frac{\kappa^2}{2} = D + c_3\lambda + c_4\frac{\lambda^2}{2}.$$

The remaining two equations were found by equating the 0th and 2nd moments from the curves to the true values 1 and 4/11 respectively. The frequency functions found were as follows:

$$\left.\begin{aligned} F(x) &= 0{\cdot}6927 - 0{\cdot}320142x^2 - 2{\cdot}7751x^3 + 3{\cdot}08x^4, \\ G(y) &= 0{\cdot}4442 - 0{\cdot}854177y + 0{\cdot}308677y^2 + 0{\cdot}649680y^3 - 0{\cdot}553667y^4, \\ H(z) &= 0{\cdot}040192z^2 + 0{\cdot}027763z^3 + 0{\cdot}008933z^4, \end{aligned}\right\} \tag{10·4}$$

where $x = t_8$.

For reasons which will be apparent in the next section, it was not deemed necessary to apply higher momental checks in this case.

Reference may here be made to yet another experiment, the negative result of which may have some interest. At the $n = 6$ stage the remarkable 'regularity' which the curve assumed, after its highly bizarre appearance at the stage before, suggested that orders of contact (except at the limit of range) might be ignored at a slightly later stage and a single curve fitted using the moments only.

Using μ_0, μ_2, μ_4 and μ_6, and the $D(\delta - z)^2/2$ (see (10·1)) for the forms at the limit of range with $F_1'(0) = 0$ the following frequency curve was found:

$$\begin{aligned} F_1(x) = {} & 0{\cdot}040192(\delta - x)^2 + 0{\cdot}132866(\delta - x)^3 - 0{\cdot}293716(\delta - x)^4 + 0{\cdot}231146(\delta - x)^5 \\ & - 0{\cdot}039279(\delta - x)^6 - 0{\cdot}005209(\delta - x)^7. \end{aligned} \tag{10·5}$$

The correct values of the moments (to 6 places) were

$$\mu_0 = 1, \quad \mu_2 = 0{\cdot}363636, \quad \mu_4 = 0{\cdot}414644, \quad \mu_6 = 0{\cdot}763334, \quad \mu_8 = 1{\cdot}823617. \tag{10·6}$$

The value μ_8' of the 8th moment computed from the curve was 1·993270, an error therefore of +9·3 %. The central ordinate $F_1(0) = 0{\cdot}9017$ as compared with the actual 0·6927, so that the curve $F_1(x)$ could not validly be used for further iteration, since the frequencies near the central frequency would be considerably in error. The probability (computed from (10·5)) for $F_1(x)$ beyond the 'true' 5 % probability point (computed from (10·3)) is 0·0456 which is quite accurate enough for practical purposes. This concordance, unexpected in view of the other facts mentioned, is due principally to the fact that $F_1(x)$ has the correct form at the limit of range. This experiment shows that, despite the regularity of the $\sqrt{b_1}$ distribution for $n = 8$, the problem of finding the nearly exact distribution cannot be treated in cavalier fashion.

11. PROBABILITY POINTS FOR FREQUENCIES FOR SAMPLES OF 8 OR MORE

By the Gram-Charlier theorem for symmetrical distributions under general conditions any frequency $f(w)$, where w has mean zero and variance unity, can be expanded in the form

$$f(w) = \exp\left\{\frac{\lambda_4}{4!\,\lambda_2^2}\left(\frac{d}{dw}\right)^4 + \frac{\lambda_6}{6!\,\lambda_2^3}\left(\frac{d}{dw}\right)^6 + \frac{\lambda_8}{8!\,\lambda_2^4}\left(\frac{d}{dw}\right)^8 + \dots\right\}\Theta(w), \tag{11·1}$$

where
$$\Theta(w) = \frac{1}{\sqrt{(2\pi)}}\exp -\tfrac{1}{2}w^2,$$

the λ being semi-invariants of the original variate. Let u be a normal variate with mean zero and variance unity. Using the method of E. A. Cornish & R. A. Fisher (1937) their expression for w in terms of u has been extended to the following effect:

$$w = u - \frac{\lambda_4}{24\lambda_2^2}x_3 - \frac{\lambda_4^2}{384\lambda_2^4}y_5 - \frac{\lambda_6}{720\lambda_2^3}x_5 + \frac{\lambda_4^3}{3072\lambda_2^6}y_7 - \frac{\lambda_4\lambda_6}{1152\lambda_2^5}z_7 - \frac{\lambda_8}{4032\lambda_2^4}x_7 + \dots, \tag{11·2}$$

where the x_k are Hermite polynomials in u of the degree indicated. The y_j and z_l terms in (11·2) are as follows:

$$\left.\begin{aligned} x_3 &= -u^3 + 3u, & y_7 &= 9u^7 - 131u^5 + 451u^3 - 321u,\\ y_5 &= 3u^5 - 24u^3 + 29u, & z_7 &= u^7 - 17u^5 + 69u^3 - 57u,\\ x_5 &= -u^5 + 10u^3 - 15u, & x_7 &= -u^7 + 21u^5 - 105u^3 + 105u.\end{aligned}\right\} \tag{11·3}$$

At (11·2) the expansion is taken to $O(n^{-3})$ because λ_{2k}/λ_2^k is $O(n^{-k+1})$ when the λ are semi-invariants of b_1 for samples of n.

The x_k, y_j and z_l functions at various probability levels are as follows:

Function	Probability points 0·10	0·05	0·025	0·01	0·001
u	1·281552	1·644854	1·959964	2·326348	3·090223
x_3	1·739867	0·484338	− 1·649229	− 5·610905	− 20·239354
y_5	− 2·97984	− 22·98240	− 37·09056	− 30·28992	+ 226·9286
x_5	− 1·632248	7·789154	16·986942	22·868797	− 33·058481
y_7	136·1309	194·9563	− 22·4505	− 675·7597	− 1286·263
z_7	19·09291	41·19959	27·20947	− 53·45639	− 261·9424
x_7	− 19·5234	− 74·2935	− 88·4883	− 15·5752	362·6625

(11·4)

For $n = 8$ the semi-invariants, etc., required are

$$\left.\begin{aligned} \lambda_2 &= 0·363636, & \lambda_4/\lambda_2^2 &= 0·1357,\\ \lambda_4 &= 0·017950, & \lambda_6/\lambda_2^3 &= -1·1612,\\ \lambda_6 &= -0·055836, & \lambda_8/\lambda_2^4 &= 2·6577.\\ \lambda_8 &= 0·046470, \end{aligned}\right\} \tag{11·5}$$

If the formula at (11·2) were quite correct and then if we computed, at any probability level ϵ, the value of w, then set $x = \lambda_2^{\frac{1}{2}}w$ and from (10·4) computed the probability from end of range the result should be exactly ϵ, assuming, of course, that (10·4) gives the exact

frequency distribution. When this procedure is carried out at different pseudo-probability, i.e. the probability of x, levels indicated, the following results are found:

Pseudo-probability	0·10	0·05	0·025	0·01	0·001
(a) True probability (to $x=\lambda_2^{\frac{1}{2}}w$)	0·096855	0·050459	0·026825	0·011504	0·001155
(b) *Normal probability* (*to* $x=\lambda_2^{\frac{1}{2}}u$)	*0·095564*	*0·052376*	*0·029502*	*0·013419*	*0·001090*

(11·6)

The correspondence at (a) is obviously satisfactory. At first sight it might appear that at 0·01 and 0·001 levels the divergence is (by the standards of this communication) rather marked. Actually this is not the case considering the fantastic difference in the algebraic form of the Gram-Charlier and the actual frequencies near the limit of range. The probabilities at (b) show that the normal curve gives quite a good representation. At $n=8$, however, the comparison flatters the normal curve since, as R. A. Fisher (1930) has shown, the ratio λ_4/λ_2^2 actually assumes its normal value of 3 at $n=7$ and reaches its greatest value at $n=22$.

We now propose to take a step which is discussed in some detail in the final section. We shall endow the right side of (11·2) with a remainder term which will make the probability of w formally the same as the pseudo-probability at (11·6). The following table shows the value of the variate $t_8\sigma^{-1}$ computed from (10·4) (where x represents t_8) at different true probability levels, together with the corresponding value of w computed from (11·2):

Probability	t_8/σ	w	R
0·10	1·253173	1·273231	−82·2
0·05	1·671682	1·666548	21·0
0·025	2·043181	2·008014	143·3
0·01	2·445125	2·389594	227·5
0·001	3·107977	3·079696	115·8

(11·7)

It has been seen that the difference between w as given by (11·2) and the true value is $O(n^{-4})$. Accordingly the values of R were found at the different probability levels by setting

$$w+\frac{R}{n^4}=\frac{t_8}{\sigma}, \tag{11·8}$$

with $n=8$. The estimates of the probability points P for values of $n \geqslant 8$ are accordingly

$$P=\lambda_2^{\frac{1}{2}}\left(A+B\frac{\lambda_4}{\lambda_2^2}+C\frac{\lambda_4^2}{\lambda_2^4}+D\frac{\lambda_6}{\lambda_2^3}+E\frac{\lambda_4^3}{\lambda_2^6}+F\frac{\lambda_4\lambda_6}{\lambda_2^5}+G\frac{\lambda_8}{\lambda_2^4}+\frac{R}{n^4}\right), \tag{11·9}$$

the values of $A, B, \ldots, G$ and R being given in the following table:

Probability	A	B	C	D	E	F	G	R
0·10	1·281552	−0·0724945	0·00776	0·00227	0·04431	−0·01657	0·000484	−82·2
0·05	1·644854	−0·0201808	0·05985	−0·01082	0·06346	−0·03576	0·001843	21·0
0·025	1·959964	0·0687179	0·09659	−0·02357	−0·00731	−0·02362	0·002195	143·3
0·01	2·326348	0·2337877	0·07888	−0·03176	−0·21997	0·04640	0·000386	227·5
0·001	3·090223	0·8433065	−0·59096	0·04591	−0·41871	0·22738	−0·008995	115·8

The terms in the first four columns agree with, or have been derived from Cornish & Fisher (1937). The λ's are semi-invariants derivable from the exact values of the moments given at (1·2).

As a test, the following is a comparison of the 0·05 and 0·01 probability points for $n = 25$ as derived by E. S. Pearson (1930) (using a Type VII curve) with the values from (11·9):

Probability level	Pearson	Geary (11·9)
0·05	0·711	0·707
0·01	1·061	1·062

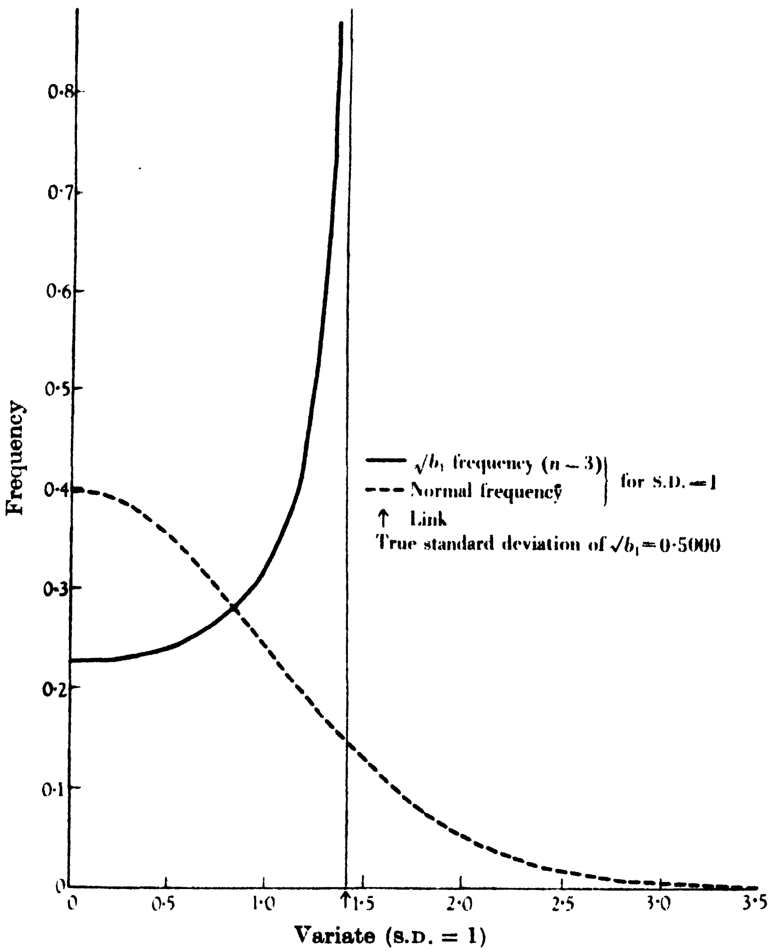


Fig. 3. Frequency of $\sqrt{b_1}$ for $n = 3$.

With standard deviation $\sigma = 0{\cdot}435$ it is obvious that the differences are not important. Sample number 25 is the lowest for which Pearson computed the probability points, and for two levels only. The formulae at (11·9) can probably be accepted with confidence.

12. Conclusion

From frequency formulae (5·2), (6·8), (7·25), (8·8), (9·10) (with (9·6) and (9·9)) and (10·4) the probability points for $\sqrt{b_1}$ for normal random samples of $n = 3, 4, 5, 6, 7$ and 8, respectively, can be determined without difficulty. The six frequency distributions are illustrated

in Figs. 3–8. On each of the $\sqrt{b_1}$ frequency curves there is superimposed the normal frequency with the same standard deviation, the intention being to enable a contrast to be made between the several $\sqrt{b_1}$ curves by reference each to the normal frequency, and to show the fairly rapid approach of the $\sqrt{b_1}$ frequency to normality with increasing n, even for small samples.*

In this research nothing was so remarkable as the transformation which the single step in the iteration, namely that from $n = 5$ to $n = 6$, effected in the shape of the frequency curve. From $n = 6$ on, the join at the links is effected so smoothly as to be almost imperceptible to the eye. The eye, however, flatters the actual approach to normality in the $\sqrt{b_1}$ frequency curves, as measured algebraically by the probability points.

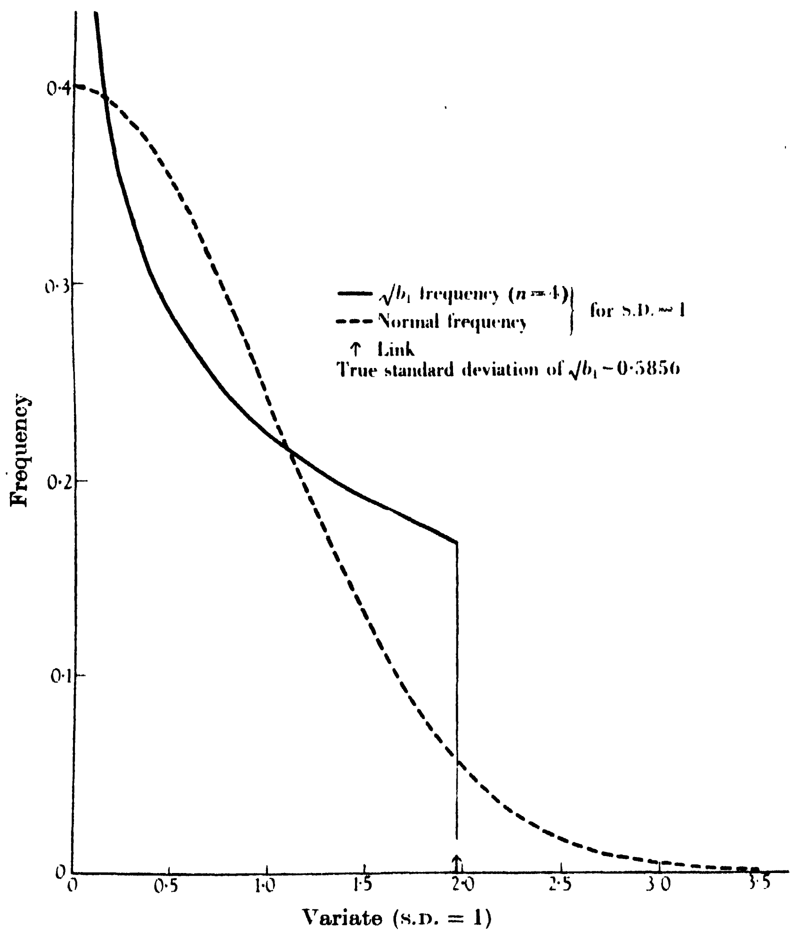


Fig. 4. Frequency of $\sqrt{b_1}$ for $n = 4$.

It may be well, at this stage, to recapitulate. Using integral iteration formula (2·13) (or (2·14)), frequency ordinates were computed at values of the variate termed the 'links' at which the frequency is shown to have functional discontinuities. Using the exact values of the moments (given at (1·2)), and taking into account the known order of contact (§ 3) of the different functions at the links and the known form assumed by the frequency at the

* As R. A. Fisher (1930) has shown, the approach to normality is not, however, uniform with increasing n, as indicated, say, by β_2. See p. 90 above.

known limit of range, inter-link frequencies were determined in polynomial form. Attention is directed to the use, at the $n = 4$ to 7 (inclusive) stages, of the higher moments for the purpose of checking the general reliability of the frequency curve (or rather series of curves joined at the links).

Of far greater practical importance, however, are the formulae (11·9) designed for the estimation of the 0·10, 0·05, 0·025, 0·01 and 0·001 probability points for normal random samples of $\sqrt{b_1}$ for $n \geq 8$. There will be little trouble about finding the corresponding formulae for other probability links. What degree of confidence can be reposed in these formulae? This raises in an acute form the vexed question (on which the protagonists of different schools were prone to get very vexed indeed a generation ago) of how best to use moments (or semi-invariants) for estimating frequency distributions. The general problem was constantly

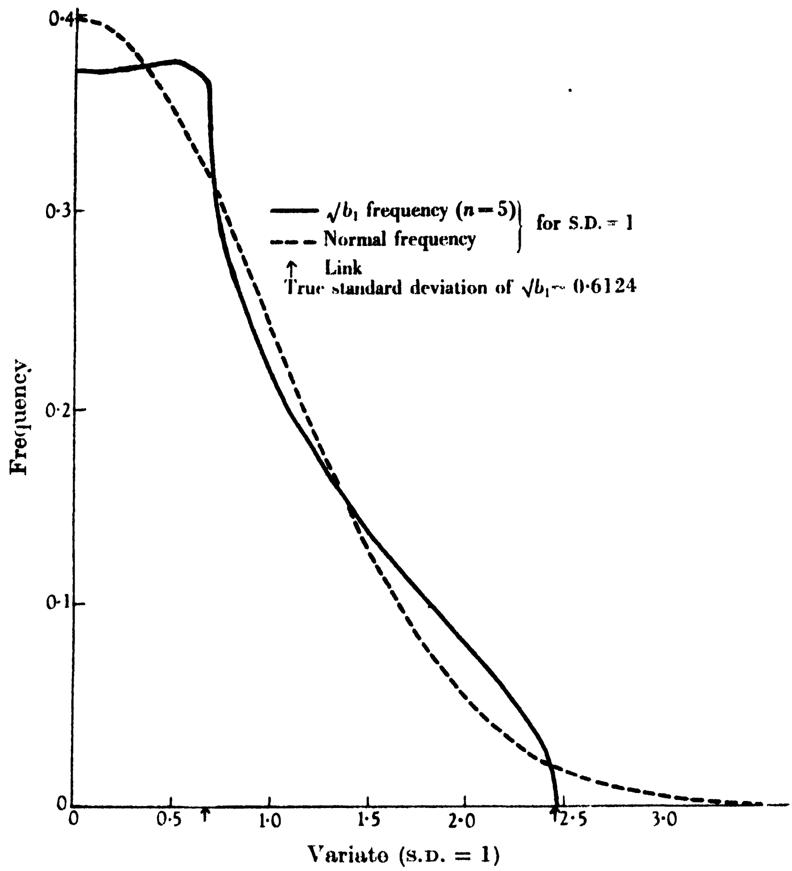


Fig. 5. Frequency of $\sqrt{b_1}$ for $n = 5$.

in the writer's mind during the present research and he would be glad if his colleagues could study the possibilities of the methods which culminated in formulae (11·9) for bridging the chasm which still divides the knowledge (sometimes exact) of the lower moments of statistics like $\sqrt{b_1}$ and b_2 and the formulae (however empirically established) for the frequency, in which a measure of confidence can be reposed. This fundamental problem was abandoned some years ago in a thoroughly unsatisfactory condition.

The Karl Pearson approach consists essentially in having regard to the 'shape' which experience has shown that frequency curves tend to assume and to use the first four moments

for the purpose of determining the constants of the curve. The disadvantage of the Pearson method is that of itself it gives no indication as to whether the resulting curve closely follows the actual frequency: it is necessary to have recourse to such devices as comparing the curve with a frequency distribution determined from hundreds of random sample computations of the statistic under examination. Apart from the tediousness of this method it is often indecisive in regard just to the parts of the frequency which are of most importance, namely the ends, because the small numbers which the check computation throws into these zones are usually subject to large (Poisson) errors.

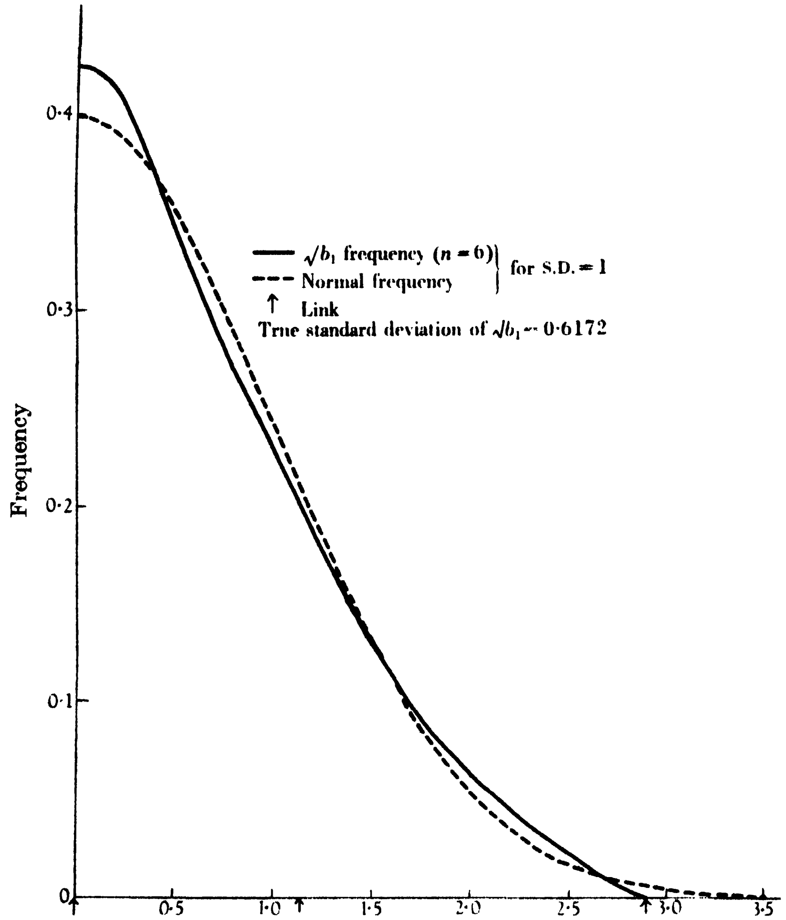


Fig. 6. Frequency of $\sqrt{b_1}$ for $n = 6$.

The Gram-Charlier system, on the other hand, can only be used with confidence when the frequency is fairly close to the normal. In practice the reliability is judged by the convergence of such terms as one can compute from the moments, i.e. if the successive terms show an 'unmistakable' tendency to diminish one feels confident in the computed frequency.

Obviously what both the Pearson, the Gram-Charlier and other frequency systems require is a Remainder Theorem. Since, however, an infinite number of moments are required to define a frequency distribution, with only a few moments known the most that can be expected is that upper (or lower) limits of the probability of the statistic can be established as

functions of the known moments. This is what Tchebychev's Theorem, and theorems of the type, do. Too much cannot be expected from the knowledge of a few moments: the approximations are almost invariably too rough for statistical use, when a high standard of efficiency is required; and M. Fréchet (1937) has shown that the Tchebychev type approximations are the best, given the assumptions, which can be made. For all their great *mathematical* importance (incidentally for their justification for the statistician of 'the faith that is in him'), it seems to the writer that research on these lines will not produce formulae which will be statistically utilizable in general conditions; but he may be quite wrong.

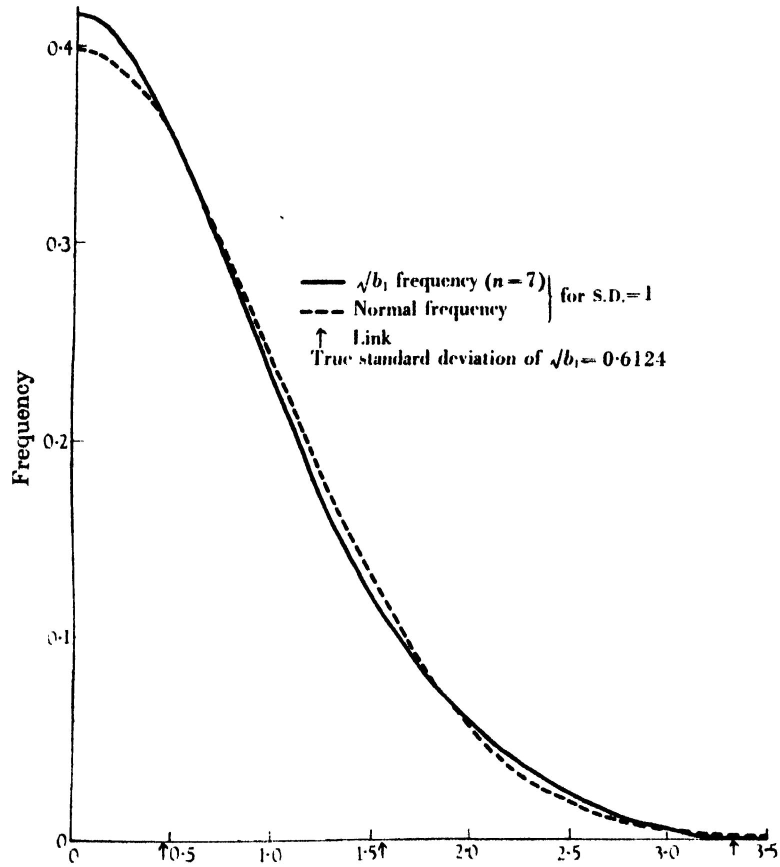


Fig. 7. Frequency of $\sqrt{b_1}$ for $n = 7$.

Knowing the earlier moments the Cornish-Fisher type expression (depending on the Gram-Charlier form of frequency) gives, at any probability level, an expansion for the variate to a defined order in the sample number. As might be surmised from the coefficients of the normal moments (e.g. (1·2) above), the coefficients in powers of $n^{-\frac{1}{2}}$ in the expansion of the variate usually tend to increase rapidly. In the present paper a remainder term of suitable order in n has been added to the known terms in the former expansion and its coefficient found by reference to the (assumed) exactly known expansion for $n = 8$. Clearly two more terms (in n^{-5} and n^{-6}) respectively could have been found had we iterated the frequency to $n = 9$ and $n = 10$, respectively, though this was not deemed necessary in the

present case. It would appear that in problems analogous to the $\sqrt{b_1}$ frequency, great precision might be obtainable by this method; it is proposed to use it with b_2; and its efficacy could be judged by iterating the frequency on a few stages further and testing the formula (with remainder) by reference to these iterations which were not used to establish the remainder.*

The iteration method is onerous but, with the co-operative effort to which it lends itself, is quite practicable and seems to have a range of application which is not confined to problems in which universal normality is assumed. Knowledge of the moments is *not* necessary for its application: in the present research the moments constituted an *embarras de choix*

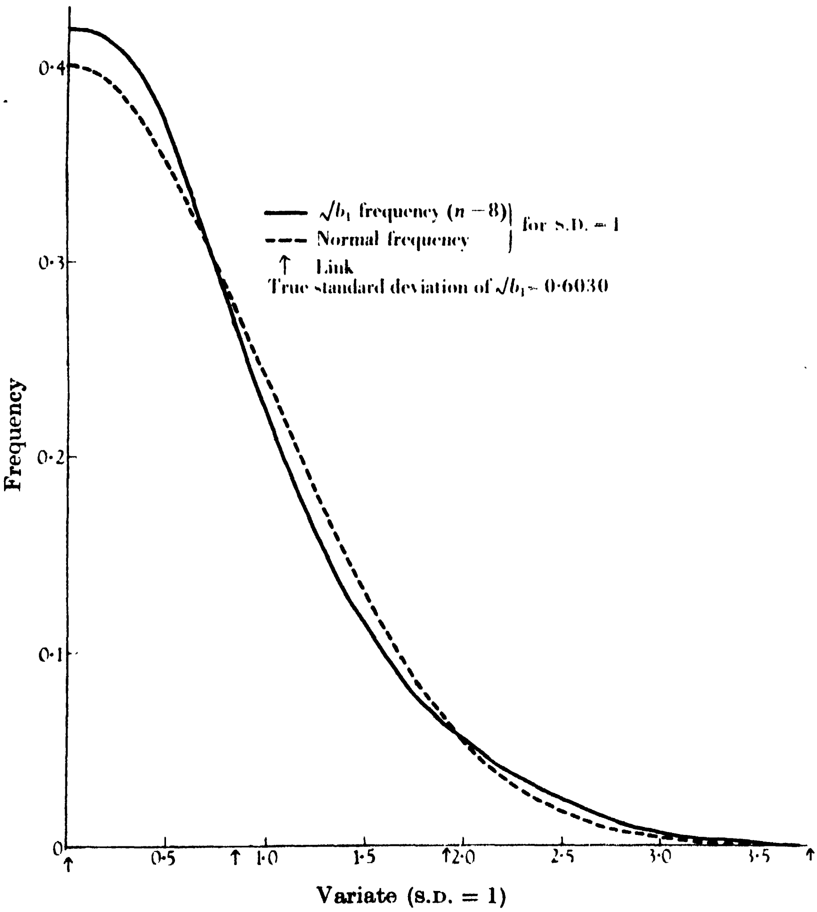


Fig. 8. Frequency of $\sqrt{b_1}$ for n = 8.

which necessitated the solution of linear equations in six or seven unknowns. In other work it may suffice to compute, from the iteration frequency, many ordinates and simply to rely on μ_0, i.e. the total frequency being unity at each stage. This is what the writer did (1935) in establishing the frequency of the test of normality a.

It is clear that the frequency formulae for $\sqrt{b_1}$ given in this communication cannot be regarded as 'proved' in the sense, say, that R. A. Fisher has proved the frequency of normal t first given by W. S. Gosset. Even for n = 4 to 8, inclusive, the iteration method yields only approximations; the method, however, can be used to attain any desired degree of precision

* Clearly the effect of the remainder diminishes rapidly with increasing n.

by increasing the number of frequencies computed by approximate integration at each stage, though this is not to be recommended to the solitary researcher. Formulae (11·9) for the probability points for $n \geq 8$ are more empirical and it would be useful to know how they fare at, say, $n = 12$, by comparison with the 'actual' frequency found by extension of the iteration, should any students be sufficiently interested in making the experiment, the results of which incidentally would be a guide in the further use of the method of integral iteration here exploited. The present research does not hold out much prospect of *exact* solutions (in the mathematical sense) of the $\sqrt{b_1}$ and b_2 frequency problems since such solutions would involve the exact knowledge of about $\frac{1}{2}n$ separate sectional frequencies in the $\sqrt{b_1}$ case. It does show that empirical solutions can be found accurately enough for practical purposes.

REFERENCES

CORNISH, E. A. & FISHER, R. A. (1937). *Revue de l'Institut international de Statistique*, 5 Année, 4 Livraison, pp. 309–20.

FISHER, R. A. (1929). *Proc. Lond. Math. Soc.* **30**, 199–238.

FISHER, R. A. (1930). *Proc. Roy. Soc.* A, **130**, 16–28.

FRÉCHET, M. (1937). *Généralités sur les Probabilités. Variables aléatoires.*

GEARY, R. C. (1935). *Biometrika*, **27**, 310–32 and 353–55.

GEARY, R. C. (1936). *Biometrika*, **28**, 295–305.

MCKAY, A. T. (1933). *Biometrika*, **25**, 204–10.

PEARSON, E. S. (1930). *Biometrika*, **22**, 239–49.

PEARSON, E. S. (1931). *Biometrika*, **22**, 423–4.

PEARSON, E. S. (1936). *Biometrika*, **28**, 306–7.

PEPPER, J. (1932). *Biometrika*, **24**, 55–64.

ON THE COMPUTATION OF UNIVERSAL MOMENTS OF TESTS OF STATISTICAL NORMALITY DERIVED FROM SAMPLES DRAWN AT RANDOM FROM A NORMAL UNIVERSE. APPLICATION TO THE CALCULATION OF THE SEVENTH MOMENT OF b_2

By R. C. GEARY and J. P. G. WORLLEDGE

1. Introductory

The principal object of this communication is to develop a computational technique appropriate to the formula given by one of the authors (Geary, 1933). By way of illustration the formula is applied to the computation of the seventh moment of

$$b_2 = \frac{m_4}{m_2^2} = n \sum_{i=1}^{n} (x_i - \bar{x})^4 / \{\Sigma (x_i - \bar{x})^2\}^2, \tag{1·1}$$

where $x_1, x_2, \ldots\ x_n$ are the measures of the random sample of n and of which $\bar{x}$ is the arithmetic mean. Universal normality is assumed throughout.

A glance at formula (3·9) in which this paper culminates will indicate that the task of deriving higher normal moments of b_2 is not one to be undertaken in a frivolous spirit. The work finds its main justification in the conviction of the authors that accurate (if not exact) values of the probability points of b_2 can be found in terms of the moments of b_2 for all values of n using a method which has proved successful in the case of the analogous test of asymmetry, involving

$$\sqrt{b_1} = \frac{m_3}{m_2^{\frac{3}{2}}} = n^{\frac{1}{2}} \Sigma (x_i - \bar{x})^3 / \{\Sigma (x_i - \bar{x})^2\}^{\frac{3}{2}}. \tag{1·2}$$

In turn, the importance of the determination of accurate probabilities for $\sqrt{b_1}$ and b_2 for normal samples derives from the facts revealed by unpublished work by one of the authors. This shows (1) that probabilistic inferences drawn from the well-known significance tests based on the assumption of universal normality are apt to go astray when, in fact, the universe is not normal, and (2) that $\sqrt{b_1}$ and b_2 provide the most efficient tests of asymmetry and kurtosis, respectively, in indefinitely large samples, amongst wide fields of alternative tests and of alternative non-normal universes.

R. A. Fisher (1930) has given the exact values of the second, fourth and sixth moments of $\sqrt{b_1}$ and J. Pepper (1932) the eighth moment. In the former paper R. A. Fisher also gave the values of the second and third moments of b_2. The moment field was extended by J. Wishart (1930) and in a joint paper by R. A. Fisher & J. Wishart (1931). C. T. Hsu & D. N. Lawley (1940) gave the fifth and sixth moments of b_2. All these authors used the combinatorial method due to R. A. Fisher (1929). The present approach is entirely different.

2. The fundamental relation

To make the *exposé* complete it may be useful to reproduce the relevant part (which is quite brief) of the 1933 paper. The method used is due essentially to C. C. Craig (1928), applied to the normal case. Using, in the usual notation, a prefixed E to indicate 'expected' or, more

accurately, 'average value for all samples', we have for the characteristic function of the

$$z_i = x_i - \bar{x} \quad (i = 1, 2, \ldots, n), \tag{2·1}$$

the expression

$$E \exp\left\{\sum_{i=1}^{n} t_i z_i\right\}, \tag{2·2}$$

where the t_i are n parameters, so that

$$E z_1^{a_1} z_2^{a_2} \ldots z_p^{a_p},$$

where the a_i are any p positive integers, is the coefficient of

$$t_1^{a_1} t_2^{a_2} \ldots t_p^{a_p}$$

in the expansion of

$$a_1!\,a_2! \ldots a_p!\, E \exp\{t_1(x_1-\bar{x}) + t_2(x_2-\bar{x}) + \ldots + t_n(x_n-\bar{x})\}. \tag{2·3}$$

The exponent can be written in the form

$$x_1(t_1-\bar{t}) + x_2(t_2-\bar{t}) + \ldots + x_n(t_n-\bar{t}),$$

where $\bar{t} = \sum_{i=1}^{n} t_i/n$. Since the x_i are independent,

$$E \exp \Sigma x_i(t_i-\bar{t}) = \prod_{i=1}^{n} E \exp x_i(t_i-\bar{t}). \tag{2·4}$$

Assuming, as we may without loss of generality, that the normal universe of the x_i has mean zero and unit standard deviation, we have

$$E \exp x_i(t_i-\bar{t}) = \exp \tfrac{1}{2}(t_i-\bar{t})^2.$$

Hence

$$a_1!\,a_2! \ldots a_p!\, E \exp \sum_{i=1}^{n} t_i(x_i-\bar{x}) = a_1!\,a_2! \ldots a_p! \exp \tfrac{1}{2} \Sigma (t_i-\bar{t})^2. \tag{2·5}$$

By definition, the *power* f of a term is given by $f = \sum_i a_i$ and the *dimension* by p. It is clear that the required universal mean value of

$$E(x_1-\bar{x})^{a_1} \ldots (x_p-\bar{x})^{a_p}$$

will be found as the coefficient of $\Pi t_i^{a_i}$ in the expansion of

$$\frac{a_1!\,a_2! \ldots a_p!}{k!\,2^k}\left\{t_1^2+t_2^2+\ldots+t_p^2 - \frac{(t_1+t_2+\ldots+t_p)^2}{n}\right\}^k, \tag{2·6}$$

where $2k = f$.

3. THE COMPUTATIONAL SCHEME

The computational scheme, which is quite general, will most clearly be outlined by reference to the computation of the exact value of a specific moment (from origin) $\mu_7'(m_4)$, for the derivation of which it was primarily designed. Then

$$\mu_7'(m_4) = \frac{1}{n^7} E(z_1^4+z_2^4+\ldots+z_n^4)^7 = \frac{1}{n^7}\Big[nE(\cdot 28 \cdot) + n(n-1)\left\{\frac{7!}{6!\,1!} E(\cdot 24 \cdot 4)\right.$$

$$\left. + \frac{7!}{5!\,2!} E(\cdot 20 \cdot 8) + \frac{7!}{4!\,3!} E(\cdot 16 \cdot 12 \cdot)\right\} + n(n-1)(n-2)\left\{\frac{7!}{5!\,1!^2\,2!} E(\cdot 20 \cdot 4^2)\right.$$

$$\left. + \frac{7!}{4!\,2!\,1!} E(\cdot 16 \cdot 8 4) + \frac{7!}{3!^2\,2!\,1!} E(\cdot 12^2 \cdot 4) + \frac{7!}{3!\,2!^2\,2!} E(\cdot 12 \cdot 8^2)\right\}$$

$$+n(n-1)(n-2)(n-3)\left\{\frac{7!}{4!\,1!^3\,3!}E(\cdot 16\cdot 4^3)+\frac{7!}{3!\,2!\,1!^2\,2!}E(\cdot 12\cdot 84^2)+\frac{7!}{2!^3\,3!\,1!}E(8^3\,4)\right\}$$

$$+n(n-1)(n-2)(n-3)(n-4)\left\{\frac{7!}{3!\,1!^4\,4!}E(\cdot 12\cdot 4^4)+\frac{7!}{2!^2\,2!\,1!^3\,3!}E(8^2 4^3)\right\}$$

$$+n(n-1)(n-2)(n-3)(n-4)(n-5)\frac{7!}{2!\,1!^5\,5!}E(84^5)$$

$$+n(n-1)(n-2)(n-3)(n-4)(n-5)(n-6)\frac{7!}{1!^7\,7!}E(4^7)\Bigg], \qquad (3\cdot 1)$$

where, for example,

$$E(\cdot 12\cdot 84^2) = Ez_1^{12}z_2^8z_3^4z_4^4 = E(x_1-\bar{x})^{12}(x_2-\bar{x})^8(x_3-\bar{x})^4(x_4-\bar{x})^4.$$

There are, accordingly, fifteen terms made up of one of dimension one, three of dimension two, four of dimension three, etc. The structure of the numerical coefficients will be noted: in particular that, when the power of a factorial appears in the denominator, its factorial also appears. Each of the fifteen E terms will be evaluated separately, grouped by dimensions and multiplied by the n-factors.

As already stated, the value of $E(a_1 a_2 a_3 \ldots)$ will be found as the coefficient of

$$t_1^{a_1}\, t_2^{a_2}\, t_3^{a_3} \ldots$$

in the expansion of
$$\frac{a_1!\,a_2!\,a_3!\ldots}{14!\,2^{14}}\{\Sigma t_i^2+\nu(\Sigma t_i)^2\}^{14}, \qquad (3\cdot 2)$$

with $\nu = -1/n$. In this case, of course, $f = \Sigma a_i = 28$.

Expand (3·2) in powers of ν by the binomial theorem. Each of the ν power terms will, in general, make a numerical contribution to the value of $E(a_1 a_2 \ldots)$ which will, accordingly, be represented by a polynomial in ν of degree 14. The term in ν^s will be

$$\frac{a_1!\,a_2!\ldots}{14!\,2^{14}}\,\frac{14!\,\nu^s}{s!(14-s)!}(14-s)!\,(2s)!\sum_s\frac{1}{(a_1-2s_1)!\,s_1!(a_2-2s_2)!\,s_2!\ldots}. \qquad (3\cdot 3)$$

In the Σ_s, summation extends to all non-negative integer series $s_1, s_2, \ldots$, so that $Ss_i = (14-s)$, s_1 being associated with a_1, s_2 with a_2, etc. The values which the s_i can assume are obviously restricted further by the condition that

$$a_i \geqslant 2s_i.$$

Let the series $(\Sigma_0, \Sigma_1, \Sigma_2, \ldots)$ be termed the *reciprocal factorial vector* (hereafter usually written 'r.f.v.') of $a_1, a_2, \ldots$, the terms of the vector being regarded as of the *order* indicated by the subscript. The vector will be indicated by clarendon type. From the computational point of view the following relation is fundamental:

$$\mathbf{A}\times\mathbf{B} = \mathbf{AB}, \qquad (3\cdot 4)$$

where $\mathbf{A} = (a_1 a_2 \ldots)$ and $\mathbf{B} = (b_1 b_2 \ldots)$, and any other r.f.v. The multiplication sign at (3·4) is defined as follows: the terms of $\mathbf{A}$ are multiplied respectively by ν^0, ν^1, ν^2, etc., and added to give a scalar A; the terms of $\mathbf{B}$ *in the reverse order* are also multiplied respectively by $\nu^0, \nu^1, \nu^2, \ldots$ and summed to give B. The coefficients of $\nu^0, \nu^1, \nu^2, \ldots$ in the product (in the ordinary sense) AB give the vector $\mathbf{AB}$. Relation (3·4) is immediately evident from the form of Σ_s in (3·3). From this relation it is quite easy to build up r.f.v.'s from those of lower order **44** from **4**, **84** from **8** and **4**, **88444** from **8** and **8444**, or **88** and **444**, etc.

Having found all fifteen r.f.v.'s the second step in the computational process is to form the scalar product of each r.f.v. and $(2s)!\,\nu^s/s!$—the latter will be termed the *ν-multipliers* which, it is important to note, are the same for all the terms in (3·1)—which, from (3·3), gives $E(a_1 a_2 \ldots)$ divided by

$$a_1!\,a_2!\ldots/2^{14}.$$

The latter are multiplied by the numerical factors in (3·1) to give what are termed the *constant multipliers*, 'constant' in the sense that they are the same for all the ν power terms in each of the E's in (3·1), but these constant terms are different for the different E's. For example, the constant multiplier for the term $Ez_1^{12}z_2^8z_3^4z_4^4$ is

$$\frac{12!\,8!\,4!^2\,7!}{3!\,2!\,1!^2\,2!\,2^{14}}. \tag{3·5}$$

Note the 'absolute constants' 7! and 2^{14}, and that the powers of the term appear as factorials in the numerator and factorials one-fourth of these powers in the denominator. In the denominator is also a 2! which is the factorial of the factorial power.

The third step in the computation is to sum the terms of the same dimensions. The final step consists in the multiplication of the terms of the different dimensions by the *ν-factors* as follows:

Table 1. *ν-factors*

Dimension	ν-factors
1	ν^6
2	$-(\nu^6+\nu^5)$
3	$2\nu^6+3\nu^5+\nu^4$
4	$-(6\nu^6+11\nu^5+6\nu^4+\nu^3)$
5	$24\nu^6+50\nu^5+35\nu^4+10\nu^3+\nu^2$
6	$-(120\nu^6+274\nu^5+225\nu^4+85\nu^3+15\nu^2+\nu)$
7	$720\nu^6+1764\nu^5+1624\nu^4+735\nu^3+175\nu^2+21\nu+1$

(3·6)

The ν-factors at (3·6) are, of course, the n-factors in (3·1) with $\nu = -1/n$.

To deal with the very large whole numbers and their reciprocals which arise in factorial computation we had recourse to a *prime number index* notation. For this purpose the number is factorized into powers of the lower primes—we have used the notation for primes not exceeding 31. Thus

$$746{,}137{,}199{,}808{,}000 = 6847\cdot 13^1\cdot 11^1\cdot 7^2\cdot 5^3\cdot 3^5\cdot 2^9$$

is written in this notation 6847[112359],

the digits in the square brackets [ ] being the powers of the lowest primes arranged in ascending order from the right. The ordinary number 6847 will be known as the *coefficient* and the symbolical number in square brackets as the *primal* of the original number. Note that in this example the notation affects an economy from 15 to 10 in the number of digits required to describe the number. Should the original number not be factorizable by a particular small prime a 0 will be inserted in the proper place, e.g. [10358] means that 7 is not a factor of the number represented. If, as often happens with the first two primes, the indices exceed 9, decimal points are used, e.g. [124·11·17] means that the original number has 2^{17} and 3^{11} as factors. The primal notation can be used when the indices are all positive

or all negative: occasionally, however, + and − signs have to be mixed in the primal (see Table 6).

With little practice great facility is acquired in applying the ordinary rules to numbers in primal notation. For multiplication or division corresponding digits in the primals are added or subtracted, the coefficients being dealt with in the ordinary way. In addition or subtraction common factors in the primals are immediately evident and the coefficient of the sum (or difference) is derived usually by a single product-sum (or product-difference) operation on a multiplying machine. It may be observed that all the work for this paper was executed without inconvenience on small hand multiplying machines with capacity $9\times 8\times 13$.

In the following tables the first thirty-two factorials, the ν-multipliers and the constant multipliers required for the computation of $\mu_7'(m_4)$ are expressed in primal notation.

Table 2. *Factorials in primal notation*

0! = 1! =	[0]	17! =	[111236·15]
2! =	[1]	18! =	[111238·16]
3! =	[11]	19! =	[1111238·16]
4! =	[13]	20! =	[1111248·18]
5! =	[113]	21! =	[1111349·18]
6! =	[124]	22! =	[1112349·19]
7! =	[1124]	23! =	[11112349·19]
8! =	[1127]	24! =	[1111234·10·22]
9! =	[1147]	25! =	[1111236·10·22]
10! =	[1248]	26! =	[1112236·10·23]
11! =	[11248]	27! =	[1112236·13·23]
12! =	[1125·10]	28! =	[1112246·13·25]
13! =	[11125·10]	29! =	[11112246·13·25]
14! =	[11225·11]	30! =	[11112247·14·26]
15! =	[11236·11]	31! =	[111112247·14·26]
16! =	[11236·15]	32! =	[111112247·14·31]

Table 3. *ν-Multipliers in factorial and primal notation*

Term in		Coefficient
ν^0	: $0!\ 0!^{-1}$ =	[0]
ν^1	: $2!\ 1!^{-1}$ =	[1]
ν^2	: $4!\ 2!^{-1}$ =	[12]
ν^3	: $6!\ 3!^{-1}$ =	[113]
ν^4	: $8!\ 4!^{-1}$ =	[1114]
ν^5	: $10!\ 5!^{-1}$ =	[1135]
ν^6	: $12!\ 6!^{-1}$ =	[11136]
ν^7	: $14!\ 7!^{-1}$ =	[111137]
ν^8	: $16!\ 8!^{-1}$ =	[111248]
ν^9	: $18!\ 9!^{-1}$ =	[1111249]
ν^{10}	: $20!\,10!^{-1}$ =	[111124·10]
ν^{11}	: $22!\,11!^{-1}$ =	[111225·11]
ν^{12}	: $24!\,12!^{-1}$ =	[11111225·12]
ν^{13}	: $26!\,13!^{-1}$ =	[11111245·13]
ν^{14}	: $28!\,14!^{-1}$ =	[11111248·14]

Table 4. *Constant multipliers in factorial and primal notation*

Required for computation of the undermentioned term in (3·1)		
$E(28)$	: $28!\,7!/7!\,2^{14}$	$= [1112246\cdot 13\cdot 11]$
$E(24\cdot 4)$	: $24!\,4!\,7!/6!\,1!\,2^{14}$	$= [1111244\cdot 11\cdot 11]$
$E(20\cdot 8)$	: $20!\,8!\,7!/5!\,2!\,2^{14}$	$= [111145\cdot 11\cdot 11]$
$E(16\cdot 12)$	: $16!\,12!\,7!/4!\,3!\,2^{14}$	$= [1246\cdot 11\cdot 11]$
$E(20\cdot 4^2)$	: $20!\,4!^2\,7!/5!\,1!^2\,2!\,2^{14}$	$= [111134\cdot 11\cdot 10]$
$E(16\cdot 84)$	: $16!\,8!\,4!\,7!/4!\,2!\,1!\,2^{14}$	$= [1145\cdot 10\cdot 11]$
$E(12^2 4)$	: $12!^2\,4!\,7!/3!^2\,2!\,1!\,2^{14}$	$= [235\cdot 11\cdot 10]$
$E(12\cdot 8^2)$	: $12!\,8!^2\,7!/3!\,2!^2\,2!\,2^{14}$	$= [145\cdot 10\cdot 10]$
$E(16\cdot 4^3)$	: $16!\,4!^3\,7!/4!\,1!^3\,3!\,2^{14}$	$= [1134\cdot 9\cdot 10]$
$E(12\cdot 84^2)$	: $12!\,8!\,4!^2\,7!/3!\,2!\,1!^2\,2!\,2^{14}$	$= [134\cdot 10\cdot 10]$
$E(8^3 4)$	: $8!^3\,4!\,7!/2!^3\,3!\,1!\,2^{14}$	$= [44\cdot 8\cdot 10]$
$E(12\cdot 4^4)$	: $12!\,4!^4\,7!/3!\,1!^4\,4!\,2^{14}$	$= [12398]$
$E(8^2 4^3)$	: $8!^2\,4!^3\,7!/2!^2\,2!\,1!^3\,3!\,2^{14}$	$= [3389]$
$E(84^5)$	: $8!\,4!^5\,7!/2!\,1!^5\,5!\,2^{14}$	$= [2188]$
$E(4^7)$	: $4!^7\,7!/1!^7\,7!\,2^{14}$	$= [77]$

The theory will be illustrated by reference to the computation of $Ez_1^8 z_2^8 z_3^4 z_4^4 z_5^4 = E(8^2 4^3)$. First the r.f.v. **88444** is found as the product **884** × **44** by setting down in equal spaces the terms of **884** and on a movable slip spaced to the former the terms of **44** in reverse:

All primals are negative

884	[27]	5 [27]	109 [39]	111 [137]	803 [112·10]	1493 [114·10]	389 [122·12]	119 [124·11]	1543 [225·15]	31 [225·15]	[225·17]
MOVABLE SLIP →		[26]	[13]	7 [13]	[1]	[2]	44				

The term in **88444** from the position illustrated is that of the 5th order, namely,

$$5[4\cdot 13]+109[\underline{4}\cdot 12]+111\cdot 7[\underline{1}4\cdot 10]+803[\underline{11}2\cdot 11]+1493[114\cdot 12]$$

$$=[114\cdot 13](5\cdot 7\cdot 5+109\cdot 7\cdot 5\cdot 2+777\cdot 7\cdot 8+803\cdot 9\cdot 4+1493\cdot 2)$$

$$=[114\cdot 13](3\cdot 27737)=27737[113\cdot 13].$$

The manner of computation is indicated: first the largest (negative) digits in each of the four positions of the primals are underlined and the underlined set is regarded as the common factor. Note how, at the final stage, the factor 3 of the coefficient reduces the primal digit from 4 to 3. From the entries in the round brackets ( ) it will be clear that, as stated above, the procedure is well adapted to the multiplying machine. The full calculation of **88444** is shown in Table 5.

The identity of the r.f.v.'s from the two factorizations of **88444** constitutes an absolute check of the work. The calculation of $E(8^2 4^3)$ required for (3·1) is completed in Table 6. In practice the figures in columns (4) and (5) of this table were derived from those in column (3), and in Tables 4 and 5 by entering the latter on two movable slips and folding opposite each entry, as required. This stage of the work was rapidly executed. The sum-product of columns (1), (2) and (5) give the value of $E(8^2 4^3)$. All the r.f.v.'s required for the calculation of the E's for (3·1) are given in the appendix.

Table 5. *Calculation of reciprocal factorial vector* **88444**

All primals are negative

Order				r.f.v. of **88444**
		(i) By **884 × 44**		
0	:	[29]	=	[29]
1	:	[2·8]+5[29]	=	7[29]
2	:	7[3·10]+5[28]+109[3·11]	=	9[·11]
3	:	[3·10]+35[3·10]+109[3·10]+111[139]	=	947[13·10]
4	:	[4·13]+5[3·10]+763[4·12]+111[138]+803[112·12]	=	1811[110·13]
5	:	5[4·13]+109[4·12]+777[14·10]+803[112·11]+1493[114·12]	=	27737[113·13]
6	:	109[5·15]+111[14·10]+5621[113·13]+1493[114·11]+389[122·14]	=	1783141[125·15]
7	:	111[15·13]+803[113·13]+1493[15·13]+389[122·13]+119[124·13]	=	20627[115·13]
8	:	803[114·16]+1493[115·13]+389[23·15]+119[124·12]+1543[225·17]	=	1772417[225·17]
9	:	1493[116·16]+389[123·15]+119[25·14]+1543[225·16]+31[225·17]	=	547889[226·17]
10	:	389[124·18]+119[125·14]+1543[126·18]+31[225·16]+[225·19]	=	151331[226·19]
11	:	119[126·17]+1543[226·18]+217[226·18]+[225·18]	=	127[223·18]
12	:	1543[227·21]+31[226·18]+[126·20]	=	2329[227·21]
13	:	31[227·21]+[226·20]	=	37[227·21]
14	:	[227·23]	=	[227·23]
		(ii) By **8844 × 4**		
0	:	[29]	=	[29]
1	:	[29]+[18]	=	7[29]
2	:	[3·11]+[18]+85[3·10]	=	9[·11]
3	:	[2·10]+85[3·10]+169[12·10]	=	947[13·10]
4	:	85[4·12]+169[12·10]+11113[104·13]	=	1811[110·13]
5	:	169[13·12]+11113[104·13]+5137[114·11]	=	27737[113·13]
6	:	11113[105·15]+5137[114·11]+22703[124·13]	=	1783141[125·15]
7	:	5137[115·13]+22703[124·13]+9341[125·13]	=	20627[115·13]
8	:	22703[125·15]+9341[125·13]+90541[225·17]	=	1772417[225·17]
9	:	9341[126·15]+90541[225·17]+2453[225·16]	=	547889[226·17]
10	:	90541[226·19]+2453[225·16]+137[126·18]	=	151331[226·19]
11	:	2453[226·18]+137[126·18]+17[226·18]	=	127[223·18]
12	:	137[127·20]+17[226·18]+[226·21]	=	2329[227·21]
13	:	17[227·20]+[226·21]	=	37[227·21]
14	:	[227·23]	=	[227·23]

Finally, the E's are multiplied by the appropriate ν-factors given in Table 1, to give the value of $E(m_4^7)$. Now R. A. Fisher (1930) (see also Geary, 1933) has shown that

$$\mu_7'(b_2) = E(b_2^7) = E(m_4^7)/E(m_2^{14}) \tag{3·7}$$

and

$$E(m_2^{14}) = (n-1)(n+1)(n+3)\dots(n+23)(n+25)/n^{14}. \tag{3·8}$$

Finally,

$$\begin{aligned}\mu_7'(b_2) = (&3^7n^{13} + 211\cdot 3^7n^{12} + 64{,}802\cdot 3^6n^{12} + 13{,}154{,}290\cdot 3^5n^{10}\\ &+ 668{,}584{,}331\cdot 3^5n^9 + 25{,}489{,}306{,}481\cdot 3^5n^8 + 74{,}020{,}784{,}452\cdot 7\cdot 3^5n^7\\ &- 72{,}634{,}851{,}124\cdot 7\cdot 5\cdot 3^6n^6 + 407{,}081{,}273{,}655\cdot 7\cdot 5\cdot 3^6n^5\\ &- 1{,}287{,}510{,}783{,}723\cdot 7\cdot 5\cdot 3^6n^4 + 2{,}526{,}463{,}322{,}982\cdot 7\cdot 5\cdot 3^6n^3\\ &- 280{,}521{,}238{,}122\cdot 11\cdot 7\cdot 5\cdot 3^6n^2 + 3{,}036{,}544{,}767\cdot 13\cdot 11\cdot 7\cdot 5^2\cdot 3^6n\\ &- 135{,}393{,}525\cdot 13\cdot 11\cdot 7^25^23^6)/(n+1)(n+3)\dots(n+23)(n+25).\end{aligned} \tag{3·9}$$

4. CORROBORATION OF FORMULAE

An integral part of the present work is the technique of check. To be of value the formulae at (3·9) and (3·10) must be absolutely correct because (1) any errors made in factorial work are fairly certain to be large and (2) the formulae are designed for use when n is small, when relatively small errors in the numerical coefficients may materially affect the results. Furthermore, it is almost impossible to avoid error (even in a joint work like the present) with so many individual calculations involving numbers astronomically large. As will appear, there is a satisfactory, though not absolute, check at the final stage; but if it reveals error it does not show where the error occurred, so that, if this were the sole check, there would be no

Table 6. *Calculation of* $E(8^2 4^3)$ *from* **88444**

Term (1)	r.f.v. 88444: Coefficient (2)	r.f.v. 88444: Primal (neg.) (3)	(3) × ν-multiplier (Table 3) (4)	(4) × constant multiplier [3389] (5)
ν^0	1	[29]	[−2−9]	[3360]
ν^1	7	[29]	[−2−8]	[3361]
ν^2	9	[·11]	[1−9]	[3390]
ν^3	947	[13·10]	[−2−7]	[3362]
ν^4	1811	[110·13]	[1−9]	[3390]
ν^5	27737	[113·13]	[−8]	[3381]
ν^6	1783141	[125·15]	[10−1−2−9]	[13260]
ν^7	20627	[115·13]	[1100−2−6]	[113363]
ν^8	1772417	[225·17]	[11−10−1−9]	[112370]
ν^9	547889	[226·17]	[111−10−2−8]	[1112361]
ν^{10}	151331	[226·19]	[1111−10−2−9]	[11112360]
ν^{11}	127	[223·18]	[1111002−7]	[111133·10·2]
ν^{12}	2329	[227·21]	[1111100−2−9]	[111113360]
ν^{13}	37	[227·21]	[1111102−2−8]	[111113561]
ν^{14}	1	[227·23]	[11111021−9]	[111113590]

alternative but to face the tedium of complete recalculation. It is essential to devise an absolute check *at each stage*. This has been done for the present technique.

The first check is the $n = 1$ (or $\nu = -1$) check. This is applicable to the E's (see (3·1)) of dimension one, two and three. It derives from the fact that

$$\left\{\Sigma t_i^2 - \frac{1}{n}(\Sigma t_i)^2\right\}^{14} \equiv \left\{(1+\nu)\,\Sigma t_i^2 + 2\nu \sum_{i>j}\sum t_i t_j\right\}^{14}. \tag{4·1}$$

It will be immediately evident from the latter that when $\nu = -1$ the following terms vanish identically:

(i) all terms of one dimension;

(ii) all terms of two dimensions except those of the type $t_1^{14} t_2^{14}$ with which we are not concerned;

(iii) all terms of three or four dimensions in which the highest power exceeds 14: the latter being the highest power which, say, t_1 can assume in the expansion of

$$2^{14}\Big(\sum_{i>j}\sum t_i t_j\Big)^{14} \nu^{14}.$$

Even in terms of three dimensions in which the highest power is less than 14, e.g. in $Ez_1^{12}z_2^{12}z_3^4$, the $\nu = -1$ test can be exploited. In fact, from (2·5) and (4·1) the required terms for $\nu = -1$ are

$$E(12^2\cdot 4) = \frac{14!}{10!\,2!^2}\,\frac{12!^2\,4!}{3!^2\,2!}\,\frac{7!}{14!\,2^{14}}\,2^{14} = 12![11124],$$

$$E(12\cdot 8^2) = \frac{14!}{6!^2\,2!\,3!}\,\frac{12!\,8!^2}{2!^2\,2!}\,\frac{7!}{14!\,2^{14}}\,2^{14} = 12![3115].$$

The sum of these two terms is 131[1236·14] which should be the sum of the four E terms of dimension three in (3·1) since $E(20\cdot 4^2)$ and $E(16\cdot 84)$ are zero (for $\nu = -1$). The checks specified in this paragraph were fully applied to the terms of one, two and three dimensions in (3·1) before multiplication by the ν-factors (Table 1).

Reciprocal factorial vectors for dimensions exceeding two were checked fully by the 'double' factorization technique exemplified in Table 5. In view of the simplicity of the two subsequent processes, namely those of the ν- and constant multipliers, this check may be taken as establishing the accuracy of the E's of dimension three or more. Reference may nevertheless be made to a check at this stage, namely that the ratios of consecutive coefficients in each E exhibit a marked regularity, if correct. Any irregularity (which in the nature of the work will usually be large) must be suspect.

Assuming the accuracy of the E's in (3·1) the final stage was checked by multiplying by the ν-factors (3·6) in two ways:

(i) by straight multiplication using the primal notation;

(ii) by taking (in (3·1))

$$= (1+\nu)(1+2\nu)\dots(1+6\nu)A_1 + \nu(1+\nu)\dots(1+5\nu)A_2 + \dots + \nu^6 A_7$$

$$= (1+\nu)\dots(1+5\nu)\{(1+6\nu)A_1 + \nu A_2\} + \dots,$$

and computing in successive stages

$$B_2 = (1+6\nu)A_1 + \nu A_2, \quad B_3 = (1+5\nu)B_2 + \nu^2 A_3, \text{ etc.}$$

The results were the same.

A satisfactory check for the final stage is that of $n = 4$. A. T. McKay (1933) has, in fact, given a formula for this value of n from which the seventh moment from zero of b_2 is found to be

$$82{,}220{,}810{,}251/5^2 11\cdot 13\cdot 17\cdot 23\cdot 29,$$

which value also transpired on substituting 4 for n in (3·9). This establishes the accuracy of all the formula except possibly the part accruing from the terms in (3·1) in

$$n(n-1)\dots(n-4), \quad n(n-1)\dots(n-5) \quad \text{and} \quad n(n-1)\dots(n-6)$$

which vanish when $n = 4$.

If it adds nothing to the check in the previous paragraph it is nevertheless of interest to observe that, for $n = 3$, the value of the seventh moment of b_2 is found to be $(\frac{3}{2})^7$ which is as

it should be since, in this case, each b_2 assumes the constant value $\frac{3}{2}$, whether the samples are normal or not.

A partial check is also afforded at the final stage by the vanishing of all coefficients of powers of ν from ν^{15} to ν^{20} inclusive.

5. Conclusion

Previous investigators in this field have all used the combinatorial technique, invented by R. A. Fisher (1929) and applied in the first instance to the cumulants, which are linear functions of the sample moments. The present writers have not had sufficient experience in working the Fisher technique to decide which method is easier to apply. It is quite likely that the Fisher method is shorter. A strong point of the present computational scheme is that it lends itself to check at every stage; and the method may appeal to students who prefer the algebraical or arithmetical to the geometrical approach. For their benefit, and also in case it may later be found necessary (in connexion with the accurate determination of the probability points of b_2 for samples of all sizes) to compute higher moments than the seventh—it is almost certain the seventh will be required—we give as an appendix an extended series of reciprocal factorial vectors. From these can be derived without difficulty (i) corresponding E's, e.g. $E(x_1-\bar{x})^8(x_2-\bar{x})^4(x_3-\bar{x})^4$, on multiplication by appropriate ν- and constant multipliers and (ii) r.f.v.'s of higher powers.

APPENDIX

A selection of reciprocal factorial vectors required for the calculation of moments of b_2 for normal samples, including all used for the calculation of the seventh moment, in primal notation

All primals are negative

Order	4	8	4^2	12	84
0	[1]	[13]	[2]	[124]	[14]
1	[1]	[12]	[1]	[114]	[4]
2	[13]	[14]	7[13]	[26]	31[26]
3		[124]	[13]	[135]	7[115]
4		[1127]	[26]	[1128]	127[1128]
5				[1148]	17[1138]
6				[1125·10]	[113·10]

Order	4^3	16	12·4	8^2	84^2
0	[3]	[1127]	[125]	[26]	[15]
1	3[3]	[1125]	[123]	[24]	[13]
2	13[5]	[137]	13[135]	5[26]	17[25]
3	3[4]	[237]	[35]	31[136]	53[125]
4	13[17]	[113·10]	47[1138]	323[1139]	2497[1138]
5	[17]	[1259]	29[1247]	[1028]	173[1137]
6	[39]	[1125·11]	157[11259]	43[124·10]	7[139]
7		[11225·11]	[11249]	[124·10]	[1049]
8		[11236·15]	[1126·13]	[224·14]	[114·13]

Order	4^4	20	16·4	12·8	$12\cdot4^2$
0	[4]	[1248]	[1128]	[137]	[126]
1	[2]	[1148]	[1028]	[37]	[26]
2	19[14]	[113·10]	11[13·10]	31[139]	101[138]
3	5[4]	[124·8]	47[1238]	7[227]	19[136]
4	49[17]	[124·11]	253[124·11]	299[123·10]	1163[1149]
5	5[16]	[135·11]	89[125·11]	73[115·10]	53[239]
6	19[38]	[1126·13]	51[1115·13]	3713[1126·12]	629[1105·11]
7	[38]	[11226·12]	1277[11226·12]	83[1126·11]	479[1125·10]
8	[4·12]	[11236·16]	229[11234·16]	1181[1236·15]	773[1126·14]
9		[111238·16]	[11135·16]	47[1237·15]	13[1126·14]
10		[1111248·18]	[11237·18]	[1237·17]	[1127·16]

Order	$8^2 4$	84^3	4^5	24	20·4
0	[27]	[16]	[5]	[1125·10]	[1249]
1	5[27]	5[16]	5[5]	[11249]	[1238]
2	109[39]	13[8]	125[17]	[125·11]	53[125·10]
3	111[137]	143[126]	35[15]	[126·11]	31[125·10]
4	803[112·10]	1399[1119]	545[28]	[224·14]	23[124·13]
5	1493[114·10]	239[1029]	23[8]	[236·12]	13[135·11]
6	389[122·12]	6943[114·11]	545[3·10]	[1137·14]	151[1136·13]
7	119[124·11]	65[104·10]	35[39]	[11236·14]	733[11236·13]
8	1543[225·15]	277[114·14]	125[4·13]	[11237·18]	1153[11237·17]
9	31[225·15]	23[115·14]	5[4·13]	[111239·17]	587[111238·16]
10	[225·17]	[115·16]	[5·15]	[1111248·19]	67[1101247·18]
11				[1112349·19]	[1111049·18]
12				[1111234·10·22]	[1111249·21]

Order	16·8	28	12^2	8^3	$8^2 4^2$
0	[113·10]	[11225·11]	[248]	[39]	[28]
1	[1129]	[11125·11]	[237]	[28]	[17]
2	13[104·11]	[1126·13]	23[249]	[·10·]	85[39]
3	43[114·11]	[1136·12]	47[259]	47[12·10]	169[129]
4	3823[224·14]	[236·15]	289[124·12]	89[101·13]	11113[104·12]
5	1507[226·12]	[238·15]	593[136·10]	281[113·11]	5137[114·10]
6	4933[1136·14]	[1237·17]	8531[1137·12]	2833[124·13]	22703[124·12]
7	28943[11236·14]	[11337·15]	193[1136·12]	11[121·13]	9341[125·12]
8	4331[1237·18]	[11248·19]	929[1236·16]	2213[224·17]	90541[225·16]
9	79[11235·17]	[111249·19]	113[1238·15]	2089[236·16]	2453[225·15]
10	43[11237·19]	[1111249·21]	37[1248·17]	71[235·18]	137[126·17]
11	37[11348·19]	[111234·10·20]	[1249·17]	[235·18]	17[226·17]
12	[11348·22]	[1111234·10·23]	[224·10·20]	[336·21]	[226·20]
13		[1112236·10·23]			
14		[1112246·13·25]			

Order	24·4	20·8	$20 \cdot 4^2$	16·12
0	[1125·11]	[125·11]	[124·10]	[124·11]
1	[1025·11]	[25·11]	[24·10]	[24·11]
2	139[1126·13]	19[124·13]	179[125·12]	187[125·13]
3	47[1126·12]	331[136·12]	87[125·11]	379[135·12]
4	79[226·15]	7001[236·15]	151[26·14]	2819[234·15]
5	31[137·15]	1489[236·15]	1501[136·14]	17161[237·15]
6	1051[1237·17]	25513[1237·17]	1609[1036·16]	104507[1237·17]
7	79[11236·15]	197[11127·15]	27487[11236·14]	30317[11237·15]
8	73[11138·19]	30211[11247·19]	13043[11137·18]	431099[11248·19]
9	59[101239·19]	168713[111249·19]	209509[111238·18]	17177[11248·19]
10	5611[1111249·21]	310841[1111249·21]	2197[1101244·20]	5513[11249·21]
11	8011[111234·10·20]	127[1111336·20]	63653[1111249·19]	1019[1134·10·20]
12	1597[1111224·10·23]	4013[111134·10·23]	1873[1111248·22]	991[1234·10·23]
13	47[1111234·10·23]	109[111135·10·23]	101[111124·10·22]	31[1235·10·23]
14	[1111234·11·25]	[111135·10·25]	[111124·10·24]	[1235·11·25]

Order	16·84	$16 \cdot 4^3$	$12^2 4$	$12 \cdot 8^2$
0	[113·11]	[112·10]	[249]	[14·10]
1	[13·11]	[12·10]	7[249]	7[14·10]
2	29[14·13]	211[113·12]	211[25·11]	73[14·12]
3	37[113·12]	659[123·11]	119[25·10]	667[25·11]
4	52139[225·15]	3671[123·14]	3821[125·13]	22697[125·14]
5	9893[216·15]	2699[115·14]	18667[136·13]	4679[116·14]
6	299297[1136·17]	29593[1115·16]	490283[1137·15]	3257831[1137·16]
7	2511043[11237·15]	106361[11215·14]	193[1133·13]	22177[1127·14]
8	360131[11235·19]	624511[11135·18]	441773[1238·17]	374281[1236·18]
9	173497[11237·19]	884393[11236·18]	24799[1238·17]	204907[1238·18]
10	589[11208·21]	14113[10236·20]	2647[1148·19]	4783[1138·20]
11	29437[11348·20]	2771[11138·19]	677[1249·18]	2251[1248·19]
12	3307[11348·23]	2579[11238·22]	2707[224·10·21]	1277[1339·22]
13	[10249·23]	23[11238·22]	23[224·10·21]	61[1349·22]
14	[11349·25]	[11239·24]	[224·11·23]	[1349·24]

Order	$12 \cdot 84^2$	$12 \cdot 4^4$	$8^3 4$	$8^2 4^3$
0	[139]	[128]	[3·10]	[29]
1	7[139]	7[128]	7[3·10]	7[29]
2	227[14·11]	47[3·10]	235[4·12]	9[·11]
3	257[23·10]	35[39]	281[13·11]	947[13·10]
4	97523[125·13]	319[110·12]	4177[112·14]	1811[110·13]
5	245[5·13]	69971[124·12]	643[111·14]	27737[113·13]
6	61219[1106·15]	3151259[1125·14]	98797[124·16]	1783141[125·15]
7	9479[1026·13]	34637[1115·12]	907[114·14]	20627[115·13]
8	12738433[1237·17]	1202651[1126·16]	37151[215·18]	1772417[225·17]
9	46223[1136·17]	115769[1126·16]	228503[236·18]	547889[226·17]
10	53593[238·19]	9841[1125·18]	50333[236·20]	151331[226·19]
11	1937[1228·18]	1823[1127·17]	391[137·19]	127[223·18]
12	1571[1238·21]	157[1028·20]	1163[336·22]	2329[227·21]
13	53[1239·21]	[1117·20]	[325·22]	37[227·21]
14	[1239·23]	[1129·22]	[337·24]	[227·23]

Order	84^5	4^7
0	[18]	[7]
1	7[18]	7[7]
2	251[2·10]	259[19]
3	1051[12·9]	77[8]
4	182843[113·12]	2107[1·11]
5	24391[103·12]	1603[1·11]
6	127741[104·14]	29771[3·13]
7	1313[4·12]	2609[3·11]
8	423697[115·16]	29771[4·15]
9	54827[115·16]	1603[3·15]
10	11455[106·18]	2107[4·17]
11	47[6·17]	77[4·16]
12	95[106·20]	259[6·19]
13	29[117·20]	7[6·19]
14	[117·22]	[7·21]

REFERENCES

CRAIG, C. C. (1928). *Metron*, **7**, 3.
FISHER, R. A. (1929). *Proc. Lond. Math. Soc.* (2), **30**, 199.
FISHER, R. A. (1930). *Proc. Roy. Soc.* A, **130**, 16.
FISHER, R. A. & WISHART, J. (1931). *Proc. Lond. Math. Soc.* (2), **33**, 195.
GEARY, R. C. (1933). *Biometrika*, **25**, 184.
HSU, C. T. & LAWLEY, D. N. (1940). *Biometrika*, **31**, 238.
McKAY, A. T. (1933). *Biometrika*, **25**, 411.
PEPPER, J. (1932). *Biometrika*, **24**, 55.
WISHART, J. (1930). *Biometrika*, **22**, 224.

THE ASYMPTOTICAL DISTRIBUTION OF RANGE IN SAMPLES FROM A NORMAL POPULATION

By G. ELFVING, *Helsingfors*

1. *Introductory.* Consider a sample of n observations, taken from an infinite normal population with the mean 0 and the standard deviation 1. Let $\mathbf{a}$ be the smallest and $\mathbf{b}$ the greatest of the observed values. Then $\mathbf{w} = \mathbf{b} - \mathbf{a}$ is the *range* of the sample.

For certain statistical purposes knowledge of the sampling distribution of range is needed. The distribution function, however, involves a rather complicated integral, whose exact calculation is, for $n > 2$, impossible. Tippett (1925), E. S. Pearson (1926, 1932) and McKay & Pearson (1933) have studied and calculated the mean, the standard deviation and the Pearson constants β_1, β_2 of the range. Fitting appropriate Pearson curves to the distribution by means of these parameters, Pearson (1932) has computed approximate percentage points for it. Later on, Hartley (1942) and Hartley & Pearson (1942) have, by numerical integration, tabulated the distribution function for $n = 2, \ldots, 20$.

As pointed out by Pearson, the distribution of range is very sensitive to departures from normality in the tails of the parental distribution. The effect of such departures becoming more perceptible for increasing n, the practical importance of the range distribution is, perhaps, small for large samples. Nevertheless, it seems to be at least of theoretical interest to investigate the *asymptotical* distribution of range for $n \to \infty$. This is the purpose of the present paper.* The results are summarized in a theorem at the end of the inquiry.

2. *The exact distribution. Transformations.* The joint-frequency function of the extremes $\mathbf{a}$, $\mathbf{b}$ reads, as well known,

$$f_{\mathbf{ab}}(a, b) = n(n-1)\,\phi(a)\,\phi(b)\,[\Phi(b) - \Phi(a)]^{n-2}\,\dagger \tag{2·1}$$

(cf. e.g. Cramér, 1945, p. 370). Let $\mathbf{u} = \frac{1}{2}(\mathbf{a} + \mathbf{b})$ denote the arithmetical mean of the extreme values of the sample. Making in (2·1) the transformation $\mathbf{a} = \mathbf{u} - \frac{1}{2}\mathbf{w}$, $\mathbf{b} = \mathbf{u} + \frac{1}{2}\mathbf{w}$ and integrating with respect to u, we find for the frequency function of the range the expression

$$f_{\mathbf{w}}(w) = n(n-1)\int_{-\infty}^{\infty} \phi(u - \tfrac{1}{2}w)\,\phi(u + \tfrac{1}{2}w)\,[\Phi(u + \tfrac{1}{2}w) - \Phi(u - \tfrac{1}{2}w)]^{n-2}\,du. \tag{2·2}$$

The object of our inquiry is the limiting form of the distribution (2·2). It proves, however, more advantageous to pass to the limit in the joint distribution of $\mathbf{a}$, $\mathbf{b}$ or $\mathbf{u}$, $\mathbf{w}$, before integrating with respect to u.

The asymptotical distribution of $\mathbf{a}$ and $\mathbf{b}$ has been investigated by Fisher & Tippett (1928), and Gumbel (1936) (cf. also Cramér, 1945, p. 376). According to these authors, we have

$$\left.\begin{aligned} E(\mathbf{u}) &= 0, & D(\mathbf{u}) &= O(\log^{-\frac{1}{2}} n), \\ E(\mathbf{w}) &= 2\sqrt{(2\log n)} + O\left(\frac{\log\log n}{\sqrt{(\log n)}}\right), & D(\mathbf{w}) &= O(\log^{-\frac{1}{2}} n). \end{aligned}\right\} \tag{2·3}$$

From the formulae quoted it is seen that $\mathbf{u} \to 0$, $\mathbf{w} \to \infty$ *in probability* as $n \to \infty$. Our first task must, consequently, be a transformation of the variables $\mathbf{a}$, $\mathbf{b}$—or $\mathbf{u}$, $\mathbf{w}$—depending on n and intended to stabilize the probability mass, in order to provide a limiting distribution.

* Prof. H. Wold has kindly directed my attention to this problem.

† $\Phi(x)$ denotes the distribution function and $\phi(x) = \Phi'(x)$ the frequency function of the normal distribution with mean at $x = 0$ and unit standard deviation.

Following the example of the authors mentioned above, we should have to introduce the new variables

$$\mathbf{a}' = n\Phi(\mathbf{a}), \quad \mathbf{b}' = n\Phi(-\mathbf{b}).$$

For our purpose it proves, however, advantageous to subject $\mathbf{a}'$ and $\mathbf{b}'$ to a new transformation, independent of n, taking

$$\left.\begin{aligned} \mathbf{x}e^{\mathbf{y}} &= 2n\Phi(\mathbf{a}) &&= 2n\Phi(-\tfrac{1}{2}\mathbf{w}+\mathbf{u}),\\ \mathbf{x}e^{-\mathbf{y}} &= 2n\Phi(-\mathbf{b}) &&= 2n\Phi(-\tfrac{1}{2}\mathbf{w}-\mathbf{u}).\end{aligned}\right\} \tag{2·4}$$

Conversely,

$$\left.\begin{aligned} \mathbf{x} &= 2n\sqrt{[\Phi(\mathbf{a})\,\Phi(-\mathbf{b})]} = 2n\sqrt{[\Phi(-\tfrac{1}{2}\mathbf{w}+\mathbf{u})\,\Phi(-\tfrac{1}{2}\mathbf{w}-\mathbf{u})]},\\ \mathbf{y} &= \tfrac{1}{2}\log\frac{\Phi(\mathbf{a})}{\Phi(-\mathbf{b})} = \tfrac{1}{2}\log\frac{\Phi(-\tfrac{1}{2}\mathbf{w}+\mathbf{u})}{\Phi(-\tfrac{1}{2}\mathbf{w}-\mathbf{u})}.\end{aligned}\right\} \tag{2·5}$$

As $\mathbf{a} \leqq \mathbf{b}$ and thus $\Phi(\mathbf{a})+\Phi(-\mathbf{b}) \leqq 1$, it follows from (2·4), that $\mathbf{x}, \mathbf{y}$ are subjected to the restrictions

$$\mathbf{x} \geqq 0, \quad \mathbf{x}\cosh\mathbf{y} \leqq n. \tag{2·6}$$

Performing the transformation, we find

$$\left|\frac{\partial(a,b)}{\partial(x,y)}\right| = \frac{x}{2n^2\phi(a)\,\phi(b)}, \tag{2·7}$$

and thus, letting $f_n(x,y)$ denote the joint-frequency function of $\mathbf{x}, \mathbf{y}$,

$$f_n(x,y) = \frac{n-1}{2n}\,x\left(1-\frac{x\cosh y}{n}\right)^{n-2}. \tag{2·8}$$

This formula is valid in the region (2·6); outside of it, we have to put $f_n(x,y) = 0$.

The new variables $\mathbf{x}, \mathbf{y}$ depend, of course, on $\mathbf{u}$ as well as $\mathbf{w}$. It will, however, be shown later, that $\mathbf{x}$, for large n, tends to coincide with the variable

$$\mathbf{x}^* = 2n\Phi(-\tfrac{1}{2}\mathbf{w}),$$

which depends exclusively on $\mathbf{w}$. For testing purposes, the former variable may thus, in large samples, be used as a substitute for the range. These considerations justify the transformation (2·4) as well as a closer study of the distribution of $\mathbf{x}$ and its limiting form.

3. *Limit passage and remainder term.* The limiting form of the joint-frequency function (2·8) is immediately seen to be

$$f(x,y) = \tfrac{1}{2}xe^{-x\cosh y} \quad (x \geqq 0). \tag{3·1}$$

The integral of this function, taken over the whole half-plane $x \geqq 0$, is easily seen to equal 1; (3·1) is, consequently, the frequency function of a well-determined two-dimensional distribution.

Let the marginal distribution functions in x, corresponding to (2·8) and (3·1), be denoted by $F_n(x)$ and $F(x)$ respectively. Our next task will be to estimate the remainder $|F_n(x)-F(x)|$, which is, obviously, at most equal to the integral

$$\Delta_n = \int_0^x\int_0^\infty 2\,|f_n(\xi,\eta)-f(\xi,\eta)|\,d\xi\,d\eta. \tag{3·2}$$

To begin with, we estimate the quotient f_n/f upwards. By differentiation with respect to the variable $z = x\cosh y$, this quotient is found to attain the maximum value

$$\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)^{n-2}e^2 = 1+\frac{1}{n}+O\left(\frac{1}{n^2}\right)$$

for $z = 2$. We thus find, for example,

$$\frac{f_n}{f} < 1+\frac{3}{2n} \quad (n \geqq 5). \tag{3·3}$$

For the further estimations, it proves necessary to divide the domain of integration in (3·2) into an interior and an exterior part by means of a convenient abscissa $\eta = y$. In order to secure the Maclaurin expansion of $\log\left(1-\frac{1}{n}\xi\cosh\eta\right)$ within the interior region, we have to choose y so as to satisfy the inequality $\frac{x\cosh y}{n} \leq k$ with an appropriate $k<1$. Taking, for simplicity, $k = 1-\sqrt{\tfrac{1}{2}}$ and observing that $\cosh y \leq e^y$, we see that the condition mentioned is fulfilled if

$$e^y \leq \frac{n}{x}(1-\sqrt{\tfrac{1}{2}}). \tag{3·4}$$

Now we may estimate f_n/f downwards in the interior domain of integration. Expanding $\log\left(1-\frac{1}{n}\xi\cosh\eta\right)$, we find

$$\log\frac{f_n}{f} = \log\left(1-\frac{1}{n}\right)+\frac{2}{n}\xi\cosh\eta-\frac{n-2}{2n^2}\xi^2\cosh^2\eta\left(1-\vartheta\frac{\xi\cosh\eta}{n}\right)^{-2} \quad (0<\vartheta<1). \tag{3·5}$$

According to the determination of y, the remainder factor is seen to be <2 for $\xi \leq x$, $\eta \leq y$. For $n \geq 3$, we have $\log\left(1-\frac{1}{n}\right) > -\frac{3}{2n}$. Omitting, further, the positive term in (3·5) and replacing $n-2$ by n, we find

$$\frac{f_n(\xi,\eta)}{f(\xi,\eta)}-1 > \log\frac{f_n(\xi,\eta)}{f(\xi,\eta)} > -\frac{\frac{3}{2}+\xi^2\cosh^2\eta}{n};$$

hence, combining with (3·3),

$$\left|\frac{f_n(\xi,\eta)}{f(\xi,\eta)}-1\right| < \frac{\frac{3}{2}+\xi^2\cosh^2\eta}{n} \quad (\xi\leq x, \eta \leq y;\ n \geq 5). \tag{3·6}$$

In the exterior domain of integration, (3·3) directly yields

$$|f_n(\xi,\eta)-f(\xi,\eta)| < f(\xi,\eta) \quad (\xi \leq x, \eta \geq y). \tag{3·7}$$

We proceed to the estimation of the integral (3·2), denoting its interior and exterior part by I_1 and I_2 respectively. For the former we have, according to (3·6), the inequality

$$I_1 = \int_0^x\int_0^y\left|\frac{f_n}{f}-1\right|2f\,d\xi\,d\eta < \frac{1}{n}\int_0^x\int_0^y(\tfrac{3}{2}\xi+\xi^3\cosh^2\eta)\,e^{-\xi\cosh\eta}\,d\xi\,d\eta, \tag{3·8}$$

for the latter, according to (3·7),

$$I_2 = \int_0^x\int_y^\infty 2\,|f_n-f|\,d\xi\,d\eta < \int_0^x\int_y^\infty \xi e^{-\xi\cosh\eta}\,d\xi\,d\eta. \tag{3·9}$$

The integration with respect to ξ may be explicitly performed. We have, in fact, putting for brevity $\cosh\eta = a$,

$$\int_0^x \xi e^{-a\xi}\,d\xi = \frac{1}{a^2}\{1-e^{-ax}[1+ax]\}, \tag{3·10}$$

$$\int_0^x \xi^3 e^{-a\xi}\,d\xi = \frac{6}{a^4}\left\{1-e^{-ax}\left[1+ax+\frac{(ax)^2}{2}+\frac{(ax)^3}{6}\right]\right\}. \tag{3·11}$$

In order to deduce remainder formulas for (a) moderate, (b) small x, we omit in (3·10) and (3·11), (a) all the negative terms, (b) the terms with x^2 and x^3. According to the Maclaurin expansion

$$e^{ax} = 1+ax+e^{\vartheta ax}\frac{(ax)^2}{2} \quad (0<\vartheta<1),$$

the expression in curled brackets in (3·10) is at most equal to $\frac{1}{2}a^2x^2$. Inserting these estimations in (3·8), we obtain for the interior integral the inequalities

$$I_1 < \frac{15}{2n}\int_0^y \frac{d\eta}{\cosh^2\eta} = \frac{15}{2n}\operatorname{tgh} y < \frac{15}{2n}, \tag{3·12a}$$

$$I_1 < \frac{15}{4n}x^2\int_0^y d\eta \leqslant \frac{4x^2}{n}y. \tag{3·12b}$$

For the exterior integral, (3·10) yields

$$I_2 < \int_y^\infty \frac{d\eta}{\cosh^2\eta} = 1 - \operatorname{tgh} y < 2e^{-2y}. \tag{3·13}$$

Finally, we have to join the results (3·12) and (3·13). Combining, first, (3·12a) with (3·13) and determining e^{-y} from (3·4) (taken with the equality sign), we obtain, after some slight simplifications in the numerical coefficients,

$$\Delta_n < \frac{8}{n}\left(1 + \frac{3x^2}{n}\right) \quad (n \geqq 5). \tag{3·14a}$$

Combining, on the other hand, (3·12b) with (3·13), we find

$$\Delta_n < \frac{4x^2}{n}y + 2e^{-2y}.$$

This expression attains, for fixed x and n, its minimum when $y = \log\frac{\sqrt{n}}{x}$. For $n \geqq 12$, this value of y also satisfies (3·4), and we obtain, as a parallel estimate to (3·14a),

$$\Delta_n < \frac{4x^2}{n}\left(\log\frac{\sqrt{n}}{x} + \frac{1}{2}\right) \quad (n \geqq 12). \tag{3·14b}$$

The formulas (3·14a, b) are both valid for all positive x and all $n \geqq 12$.

4. *The asymptotical distribution.* Having established the limiting distribution of the variable $\mathbf{x}$ defined in (2·5), we are going to examine its properties.

The frequency function of the distribution considered reads, according to (3·1),

$$f(x) = x\int_0^\infty e^{-x\cosh y}\,dy = x\int_1^\infty \frac{e^{-xt}}{\sqrt{(t^2-1)}}\,dt. \tag{4·1}$$

Changing the order of integration, we easily find the distribution function, the mean and the variance of (4·1) to be

$$F(x) = 1 - \int_0^\infty \frac{1 + x\cosh y}{\cosh^2 y}e^{-x\cosh y}\,dy = 1 - \int_1^\infty \frac{1+xt}{t^2\sqrt{(t^2-1)}}e^{-xt}\,dt, \tag{4·1'}$$

$$E(\mathbf{x}) = \tfrac{1}{2}\pi, \quad D^2(\mathbf{x}) = 4 - \tfrac{1}{4}\pi^2. \tag{4·2}$$

The numerical evaluation of the distribution is much simplified by the fact that $f(x)$ as well as $F(x)$ is closely connected with certain *Bessel functions*. Denote

$$\phi(x) = \int_0^\infty e^{-x\cosh y}\,dy = \int_1^\infty \frac{e^{-xt}}{\sqrt{(t^2-1)}}\,dt. \tag{4·3}$$

By differentiation and partial integration, this function is found to satisfy the differential equation

$$\phi''(x) + \frac{1}{x}\phi'(x) - \phi(x) = 0. \tag{4·4}$$

Changing x into $-ix$, we obtain for the function $\psi(x) = \phi(-ix)$ the equation

$$\psi''(x)+\frac{1}{x}\psi'(x)+\psi(x) = 0; \tag{4·4'}$$

hence, $\psi(x)$ is a Bessel function of order zero.

In order to specify this function, we will deduce an asymptotical expression for the function (4·3), valid for large x. For this purpose, we make in the latter integral (4·3) the substitution $t = 1+u/x$ and write

$$\left(1+\frac{u}{2x}\right)^{-\frac{1}{2}} = 1-\vartheta\frac{u}{4x} \quad (0<\vartheta<1).$$

Performing the integration, we obtain

$$\phi(x) = \sqrt{\left(\frac{\pi}{2}\right)}\, x^{-\frac{1}{2}}\, e^{-x}\left(1+O\left(\frac{1}{x}\right)\right), \tag{4·5}$$

which shows that the Bessel function $\psi(x) = \phi(-ix)$ tends to zero for $x\to+i\infty$. This function is, consequently, proportional to the *Hankel function* $H_0^{(1)}(x)$ (cf. Jahnke-Emde, 1909, p. 94). Comparing the asymptotical expressions of $\phi(x)$ and $iH_0^{(1)}(ix)$, we find the proportional factor to be $\frac{1}{2}\pi$, whence

$$f(x) = x\frac{\pi i}{2}H_0^{(1)}(ix). \tag{4·6}$$

We proceed to the calculation of $F(x)$. Every integral of $xH_0^{(1)}(x)$ is (cf. Jahnke-Emde, p. 165) of the form $xH_1^{(1)}(x)+\text{Const.}$, where $H_1^{(1)}(x)$ is the *first* order Hankel function corresponding to $H_0^{(1)}(x)$; consequently,

$$F(x) = \frac{\pi x}{2}H_1^{(1)}(ix)+C.$$

Now $\frac{\pi x}{2}H_1^{(1)}(ix)$ tends to zero as $-(\frac{1}{2}\pi x)^{\frac{1}{2}}e^{-x}$ for $x\to\infty$ (cf. Jahnke-Emde, 1909, p. 101); hence $C = 1$ and

$$F(x) = 1-x\left[-\frac{\pi}{2}H_1^{(1)}(ix)\right]. \tag{4·7}$$

For small x, $F(x)$ has the expansion

$$F(x) = \left(\log\frac{2}{\gamma x}+\frac{1}{2}\right)\frac{x^2}{2}+\left(\log\frac{2}{\gamma x}+\frac{5}{4}\right)\frac{x^4}{16}+\dots, \tag{4·8}$$

where

$$\log\frac{2}{\gamma} = 0·11593.... \tag{4·9}$$

The factors of x in (4·6) and (4·7) are tabulated in Jahnke-Emde (1909, pp. 135–6). Below, we give a short table of $f(x)$ and $F(x)$. The corresponding curves are seen in Fig. 1.

5. *Connexion between the variable* $\mathbf{x}$ *and the range.* We now turn back to the original object of our inquiry: the asymptotical distribution of the range.

Consider the variable $\mathbf{x} = 2n\sqrt{[\Phi(-\frac{1}{2}\mathbf{w}+\mathbf{u})\,\Phi(-\frac{1}{2}\mathbf{w}-\mathbf{u})]}$ (5·1)

introduced in (2·4). As mentioned earlier,

$$\mathbf{w}\to\infty, \quad \mathbf{u}\to 0 \text{ in probability} \quad (n\to\infty). \tag{5·2}$$

Under such circumstances, for large n, $\mathbf{x}$ may be expected to behave substantially as the variable

$$\mathbf{x}^* = 2n\Phi(-\tfrac{1}{2}\mathbf{w}), \tag{5·3}$$

which depends *exclusively on the range.*

We shall now prove that $\mathbf{x}^*/\mathbf{x} \to 1$ in probability as $n \to \infty$. According to the well-known asymptotic formula

$$\Phi(-x) = \frac{1}{x\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}\left(1 - \frac{\vartheta}{x^2}\right) \quad (x > 0); \quad 0\,(< \vartheta < 1),$$

we may, for $|\mathbf{u}| < \frac{1}{2}\mathbf{w}$, write

$$\frac{\mathbf{x}^*}{\mathbf{x}} = e^{\frac{1}{2}\mathbf{u}^2}\left(1 - \frac{4\mathbf{u}^2}{\mathbf{w}^2}\right)^{\frac{1}{2}} \{1 + O[(\tfrac{1}{2}\mathbf{w} - |\mathbf{u}|)^{-2}]\}.$$

x	$f(x)$	$F(x)$	x	$f(x)$	$F(x)$
0·0	0·0000	0·0000	1·5	0·3207	0·5839
0·1	0·2427	0·0146	2·0	0·2278	0·7202
0·2	0·3505	0·0448	2·5	0·1559	0·8153
0·3	0·4118	0·0832	3·0	0·1042	0·8795
0·4	0·4458	0·1262	4·0	0·0446	0·9501
0·5	0·4622	0·1718	5·0	0·0185	0·9798
0·6	0·4665	0·2183	6·0	0·0075	0·9919
0·7	0·4624	0·2648	7·0	0·0030	0·9968
0·8	0·4522	0·3106	8·0	0·0012	0·9988
0·9	0·4380	0·3552	9·0	0·0005	0·9995
1·0	0·4210	0·3981	10·0	0·0002	0·9998

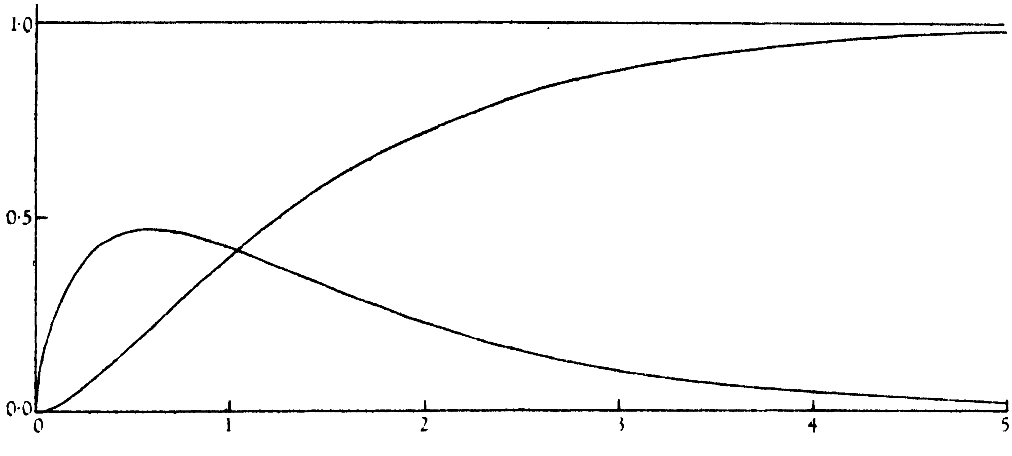


Fig. 1

Given an arbitrary $\epsilon > 0$, we obviously may find two positive numbers u_ϵ and w_ϵ $(> u_\epsilon)$ such that

$$\left|\frac{\mathbf{x}^*}{\mathbf{x}} - 1\right| < \epsilon \quad \text{if} \quad \mathbf{w} \geqq w_\epsilon, \quad |\mathbf{u}| \leqq u_\epsilon. \tag{5·4}$$

On account of (5·2), we may, on the other hand, choose n_ϵ so that the probability of the simultaneous validity of the latter inequalities in (5·4) exceeds $1 - \epsilon$ if $n \geqq n_\epsilon$. Consequently,

$$P\left\{\left|\frac{\mathbf{x}^*}{\mathbf{x}} - 1\right| < \epsilon\right\} > 1 - \epsilon \quad (n \geqq n_\epsilon), \tag{5·5}$$

which proves our statement.

As shown in section 3, the distribution function $F_n(x)$ of $\mathbf{x}$ converges to $F(x)$ as $n \to \infty$. Since $F(0) = 0$, it follows from (5·5), by a well-known method of argument, that the distribution function $F_n^*(x)$ of $\mathbf{x}^*$ converges to the same limiting function. The asymptotical distribution of the range, suitably transformed, is hereby established.

For practical purposes, it would, of course, be desirable to possess a reasonably accurate estimate of the remainder $F_n^*(x) - F(x)$, or at least an estimate of the difference $F_n^*(x) - F_n(x)$, to be combined with the results (3·14).

For $n = 20$, the accuracy of $F(x)$ as substitute for $F_n^*(x)$ may be checked by means of Hartley's (1942) tables. The discrepancy amounts to about 0·004 for $x = 0{\cdot}1$, 0·025 for $x = 1$ and 0·010 for $x = 4$.

The theoretical evaluation of $F_n^*(x) - F(x)$ seems to be somewhat complicated and, besides, of little use since $\mathbf{x}^*$, for most purposes, may be replaced by $\mathbf{x}$. A few remarks concerning the relations between $\mathbf{x}$, $\mathbf{x}^*$ and their distribution functions will, however, be added below.

To begin with, we note that always $\mathbf{x} \leqq \mathbf{x}^*$, the equality sign being valid only if $\mathbf{u} = 0$. Consider, in fact, the function $x(u)$, defined by (5·1) for a fixed w. Inserting for Φ its analytical expression, we easily find that $D^{(2)} \log x(u) \leqq 0$ for all u. Hence, $x(u)$ has no minimum and at most one maximum, and the latter is, by symmetry, seen to be attained for $u = 0$, being thus equal to $\mathbf{x}^*$.

From $\mathbf{x} \leqq \mathbf{x}^*$, it follows that $F_n^*(x) \leqq F_n(x)$ for all x. We will show that the difference $F_n(x) - F_n^*(x)$ may be expressed as a double integral.

The variables $\mathbf{u}$ and $\mathbf{w}$ are, according to (2·4), well-determined functions of $\mathbf{x}$ and $\mathbf{y}$ in the region (2·6); and so is the variable $\mathbf{x}^*$, on account of (5·3).

On the *level curve* $\mathbf{x}^* = x_0$, $\mathbf{w}$ has a constant value w_0, determined by

$$2n\Phi(-\tfrac{1}{2}w_0) = x_0, \tag{5·6}$$

and this curve is, consequently, given in parametric form by the equations

$$x = 2n\sqrt{[\Phi(-\tfrac{1}{2}w_0 + u)\,\Phi(-\tfrac{1}{2}w_0 - u)]}, \quad y = \tfrac{1}{2}\log\frac{\Phi(-\frac{1}{2}w_0 + u)}{\Phi(-\frac{1}{2}w_0 - u)}, \tag{5·7}$$

where u runs through all values from $-\infty$ to $+\infty$. The latter function (5·7) being, obviously, monotonously increasing, we may imagine u eliminated, writing (5·7) in the form

$$x = \xi_n(x_0, y) \quad (-\infty < y < \infty). \tag{5·7'}$$

From the proof of the inequality $\mathbf{x} \leqq \mathbf{x}^*$ given above, it follows that the function (5·7') has a single maximum for $y = 0$. When $y \to \pm\infty$, the function obviously tends to zero.

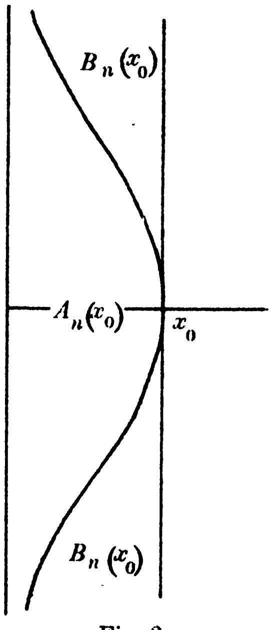


Fig. 2

The inequality $\mathbf{x}^* \leqq x_0$ is fulfilled on the left side of the curve (5·7'), the inequality $\mathbf{x} \leqq x_0$ on the left side of the straight line $\mathbf{x} = x_0$. Let us for brevity denote the regions (cf. fig. 2)

$$0 \leqq \mathbf{x} \leqq \xi_n(x_0, \mathbf{y}), \quad \xi_n(x_0, \mathbf{y}) < \mathbf{x} \leqq x_0 \tag{5·8}$$

by $A_n(x_0)$ and $B_n(x_0)$ respectively. The difference $F_n(x_0) - F_n^*(x_0)$ is, then, the probability of the points $\mathbf{x}$, $\mathbf{y}$ falling within the region $B_n(x_0)$. Dropping the indices 0, we thus obtain the expression sought for

$$F_n(x) - F_n^*(x) = \iint_{B_n(x)} f_n(\xi, \eta)\, d\xi\, d\eta. \tag{5·9}$$

Comparing, finally, the transformed range distribution function $F_n^*(x)$ directly with its limiting form $F(x)$, we find

$$\begin{aligned} F_n^*(x) - F(x) &= [F_n(x) - F(x)] - [F_n(x) - F_n^*(x)] \\ &= \iint_{\xi \leqq x} (f_n - f)\, d\xi\, d\eta - \iint_{B_n(x)} f_n\, d\xi\, d\eta \\ &= \iint_{A_n(x)} (f_n - f)\, d\xi\, d\eta - \iint_{B_n(x)} f\, d\xi\, d\eta. \end{aligned} \tag{5·10}$$

The former integral is, obviously, at most equal to the remainder expression Δ_n in (3·2), estimated in (3·14).

6. *Conclusion.* Our main results may be summarized in the following theorem:

THEOREM. Consider a sample of n observations from an infinite normal population with mean 0 and standard deviation 1. Let $\mathbf{a}$ be the smallest, $\mathbf{b}$ the greatest of the observed values, and put

$$\mathbf{x} = 2n\sqrt{[\Phi(\mathbf{a})\,\Phi(-\mathbf{b})]}, \quad \mathbf{x}^* = 2n\Phi\left(-\frac{\mathbf{b}-\mathbf{a}}{2}\right),$$

the latter variable being evidently a simple transformation of the range of the sample. Then

(1) $\mathbf{x} \leqq \mathbf{x}^*$; $\mathbf{x}^*/\mathbf{x} \to 1$ in probability ($n \to \infty$).

(2) The distribution functions $F_n(x)$ and $F_n^*(x)$ of $\mathbf{x}$ and $\mathbf{x}^*$ tend, for $n \to \infty$, to the common limit

$$F(x) = 1 - \int_1^\infty \frac{1 + xt}{t^2\sqrt{(t^2-1)}}\, e^{-xt}\, dt = 1 + \frac{\pi x}{2} H_1^{(1)}(ix),$$

where $H_1^{(1)}(z)$ is the first order Bessel function, which vanishes as $-\left(\frac{\pi z}{2i}\right)^{-\frac{1}{2}} e^{iz}$ for $z \to +i\infty$.

(3) For $n \geqq 12$, $F_n(x)$ satisfies the inequalities

$$|F_n(x) - F(x)| < \frac{8}{n}\left(1 + \frac{3x^2}{n}\right), \quad |F_n(x) - F(x)| < \frac{4x^2}{n}\left(\log\frac{\sqrt{n}}{x} + \frac{1}{2}\right).$$

7. *Generalization.* A great part of our conclusions does not presuppose the normality of the parental population. Thus, the distribution (2·8) of the variables $\mathbf{x}$, $\mathbf{y}$ defined by (2·5) is the same for any continuous probability law and so, consequently, is its limiting form; however, if the parental distribution is non-symmetrical, with distribution function $G(x)$, say, the factor $\Phi(-\mathbf{b})$ in (2·5) must, of course, be replaced by $1 - G(\mathbf{b})$ instead of $G(-\mathbf{b})$, and the variable $\mathbf{x}^*$ is to be defined by

$$\mathbf{x}^* = 2n\sqrt{\{G(-\tfrac{1}{2}\mathbf{w})\,[1 - G(\tfrac{1}{2}\mathbf{w})]\}}.$$

The proof of the statement $\mathbf{x}^*/\mathbf{x} \to 1$ requires, however, convenient assumptions concerning the parental distribution. It can be proved that the assertion mentioned—and, consequently, the theorem stated above—are valid if the frequency function of this distribution is of the form

$$g(x) = C\exp\left[-\frac{1}{p}|x|^p\right],$$

where $1 < p \leqq 2$.

REFERENCES

CRAMÉR, H. (1945). *Mathematical Methods of Statistics*. Uppsala.

FISHER, R. A. & TIPPETT, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Camb. Phil. Soc.* **24**, 180.

GUMBEL, E. J. (1936). Les valeurs extrêmes des distribution statistiques. *Ann. Inst. Poincaré*, **5**, 115–58.

HARTLEY, H. O. (1942). The range in random samples. *Biometrika*, **32**, 334–48.

HARTLEY, H. O. & PEARSON, E. S. (1942). The probability integral of the range in samples of n observations from a normal population. *Biometrika*, **32**, 301–10.

JAHNKE, E. & EMDE, F. (1909). *Funktionentafeln*. Leipzig and Berlin.

MCKAY, A. T. & PEARSON, E. S. (1933). A note on the distribution of range in samples of n. *Biometrika*, **25**, 415–20.

PEARSON, E. S. (1926). A further note on the distribution of range in samples, taken from a normal population. *Biometrika*, **18**, 173–94.

PEARSON, E. S. (1932). The percentage limits for the distribution of range in samples from a normal population. *Biometrika*, **24**, 404–17.

TIPPETT, L. H. C. (1925). On the extreme individuals and the range of samples taken from a normal population. *Biometrika*, **17**, 364–87.

LIMITS OF THE RATIO OF MEAN RANGE TO STANDARD DEVIATION*

By R. L. PLACKETT, B.A.

The ratio of mean range $\bar{w}_n$ in samples of n to population standard deviation σ, which has been denoted by d_n, is used in control chart work (when the population is assumed normal) to estimate σ from the ranges of a set of small samples. On comparing the series of values of d_n for different n when the parent population is rectangular with the series when it is normal (see table below), it is clear that for $n \leqslant 12$ the two series agree to within less than 10%. With this in mind, the question arises: what are the limiting values of d_n for a given n? It is shown here that populations exist for which d_n is arbitrarily near to zero, while for no population will d_n exceed the value

$$n\sqrt{\left(\frac{2}{(2n-1)!}\{(2n-2)!-[(n-1)!]^2\}\right)}.$$

We consider a population whose distribution function is $F(x)$ and which extends from $-a$ to $+a$ so that $F(-a)=0$ and $F(a)=1$. The population in the first place may have any finite limits, but there is no loss in generality in supposing these. It is required to find limits to the ratio

$$d_n=\frac{\int_{-a}^{a}[1-F^n-(1-F)^n]\,dx}{\left[\int_{-a}^{a}x^2\,dF-\left(\int_{-a}^{a}x\,dF\right)^2\right]^{\frac{1}{2}}}. \tag{1}$$

We apply the calculus of variations and find the extremes of d_n in the class of functions F such that $F(-a)=0$ and $F(a)=1$; the case is thus one of fixed end-points. Suppose that $F(x)=u(x)$ gives an extreme value and form the functions $F(x)=u(x)+tv(x)$; for t suitably near to zero, all these will be permissible distribution functions, i.e. monotonically increasing, provided $v(-a)=v(a)=0$. Then for $t=0$, $d/dt(d_n)$ is zero for all functions $v(x)$.

Since
$$d_n=\frac{\int_{-a}^{a}[1-(u+tv)^n-(1-u-tv)^n]\,dx}{\left[\int_{-a}^{a}x^2(u'+tv')\,dx-\left(\int_{-a}^{a}x(u'+tv')\,dx\right)^2\right]^{\frac{1}{2}}},$$

$$\left[\frac{d}{dt}(d_n)\right]_{t=0}=\frac{\left[\begin{array}{l}2n\left[\int_{-a}^{a}x^2u'dx-\left(\int_{-a}^{a}xu'dx\right)^2\right]\left[\int_{-a}^{a}(1-u)^{n-1}v\,dx-\int_{-a}^{a}u^{n-1}v\,dx\right]\\ \quad-\left[\int_{-a}^{a}(1-u^n-(1-u)^n)\,dx\right]\left[\int_{-a}^{a}x^2v'dx-2\left(\int_{-a}^{a}xu'dx\right)\left(\int_{-a}^{a}xv'dx\right)\right]\end{array}\right]}{2\left[\int_{-a}^{a}x^2u'dx-\left(\int_{-a}^{a}xu'dx\right)^2\right]^{\frac{3}{2}}}$$
$$=0.$$

Now $\int_{-a}^{a}x^2v'dx=-2\int_{-a}^{a}xv\,dx$ since $v(a)=v(-a)=0$, and by the same condition

$$\int_{-a}^{a}xv'dx=-\int_{-a}^{a}v\,dx.$$

* Communication from the National Physical Laboratory.

The numerator now becomes of the form $\int_{-a}^{a} s(x)\,v(x)\,dx$, and this must be zero for all functions $v(x)$; it is therefore concluded that $s(x)$ is identically equal to zero. In fact

$$n\left[\int_{-a}^{a} x^2u'dx-\left(\int_{-a}^{a} xu'dx\right)^2\right][(1-u)^{n-1}-u^{n-1}]$$
$$=\left[\int_{-a}^{a}\{1-u^n-(1-u)^n\}\,dx\right]\left[\int_{-a}^{a} xu'dx-x\right],$$

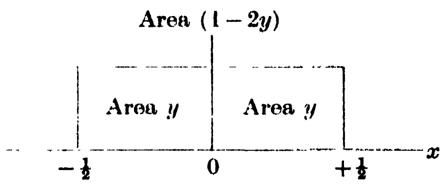
so that if μ is the mean, σ the standard deviation, $\overline{w}_n$ the mean range in samples of n, and $F(x)$ the distribution function of the population which gives an extreme value to d_n, we have

$$\overline{w}_n(x-\mu)=n\sigma^2[F^{n-1}-(1-F)^{n-1}].$$

Put $x=-a$ and obtain $n\sigma^2=\overline{w}_n(\mu+a)$; $x=a$ gives $n\sigma^2=\overline{w}_n(a-\mu)$ whence $\mu=0$ and

$$x=a[F^{n-1}-(1-F)^{n-1}]. \tag{2}$$

This distribution must give an upper limit to d_n since if we consider a distribution of the type below:



$F(x)=(2x+1)\,y \qquad -\frac{1}{2}\leqslant x<0$

$F(x)=1-(1-2x)\,y \qquad 0<x\leqslant\frac{1}{2}$

$y\leqslant\frac{1}{2}$

the ratio (1) for $y=O(n^{-3})$ is approximately $\sqrt{(3/2)}\,n\sqrt{y}$ which can be made as small as we please.

Reverting therefore to equation (2) we note that since $a\overline{w}_n=n\sigma^2$, $d_n(\text{max.})=n\sigma/a$.

$$\sigma^2=\int_{-a}^{a}x^2dF-\left(\int_{-a}^{a}x\,dF\right)^2$$
$$=a^2\int_{-a}^{a}[F^{n-1}-(1-F)^{n-1}]^2\,dF.$$

Therefore
$$\sigma^2/a^2=\frac{2}{2n-1}-2B(n,n),$$

i.e.
$$d_n\,(\text{max.})=n\sqrt{\left(\frac{2}{(2n-1)!}\{(2n-2)!-[(n-1)!]^2\}\right)}. \tag{3}$$

It is of interest to note that all the foregoing analysis may be carried out with a equal to any finite value and so we may take the limit as $a\to\infty$, and equation (3), which is independent of a, will still hold.

It is easy to verify, by Stirling's formula or otherwise, that as n increases $[(n-1)!]^2$ becomes negligible compared with $(2n-2)!$.

Consequently, for large n,
$$d_n(\text{max.})\simeq n\sqrt{\{2/(2n-1)\}}$$
$$=\sqrt{\left(n+\frac{1}{2-1/n}\right)},$$

i.e.
$$d_n(\text{max.})\simeq\sqrt{(n+\tfrac{1}{2})}. \tag{4}$$

The probability density function of (2) is obtained by differentiation and is

$$f(x)=\frac{1}{a(n-1)\,[F^{n-2}+(1-F)^{n-2}]}, \tag{5}$$

so that (2) and (5) are the parametric equations of the curve in terms of its distribution function. Thus for $n > 2$, $f(0) = \frac{2^{n-3}}{a(n-1)}$ and $f(\pm a) = \frac{1}{a(n-1)}$. The distributions (2) are readily seen to be unimodal and symmetrical about $x = 0$. For $n = 2, 3$ they are rectangular. For $F > \frac{1}{2}$ and large n, $aF^{n-1} \simeq x$. Hence $F^{n-2} \simeq \frac{x}{a}$, $f(x) \simeq \frac{1}{x(n-1)}$. Similar considerations for $F < \frac{1}{2}$ show that for large n and $x \neq 0$,

$$f(x) \simeq \frac{1}{|x|(n-1)}.$$

From (4), $\sigma \sim a/\sqrt{n}$. Consequently, for any finite a, as $n \to \infty$ the distributions (2) tend to a single ordinate at $x = 0$. This should be compared with the limiting case giving $d_n \to 0$ for fixed n illustrated with the diagram above. The limiting form of the two distributions is the same but the approach to the limit with increasing n is quite different. There is no approach to normality.

Following is a table of d_n(max.) and of d_n in samples from normal and rectangular populations for $n = 2, \ldots, 12$. The quantity $\sqrt{(n+\frac{1}{2})}$ is also included to see how closely (4) is approximated. The values of d_n (normal) are obtained from the paper by E. S. Pearson (1942). For a rectangular distribution d_n is simply $2\sqrt{3}(n-1)/(n+1)$.

n	$\sqrt{(n+\frac{1}{2})}$	d_n (max.)	d_n (normal)	d_n (rectangular)
2	1·58114	1·15470	1·128	1·15470
3	1·87083	1·73205	1·693	1·73205
4	2·12132	2·08395	2·059	2·07846
5	2·34521	2·34013	2·326	2·30940
6	2·54951	2·55333	2·534	2·47436
7	2·73861	2·74414	2·704	2·59808
8	2·91548	2·92076	2·847	2·69430
9	3·08221	3·08685	2·970	2·77128
10	3·24037	3·24440	3·078	2·83426
11	3·39116	3·39466	3·173	2·88675
12	3·53553	3·53860	3·258	2·93116

Some values of d_n for a number of symmetrical populations were given by Pearson & Adyanthaya (1928) and have been reproduced with some figures for one skew population in *Tables for Statisticians and Biometricians*, Part II, Table XXIII. The majority of these values were obtained empirically from random sampling experiments. These values were of course subject to sampling error and for this reason are in three cases very slightly above d_n(max.).

Some of the preceding work was done as part of the Research and Development programme of the Ministry of Supply (S.R. 17) and appears by permission of the Chief Scientific Officer. It was completed as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory.

REFERENCES

PEARSON, E. S. (1942). The probability integral of the range in samples of n observations from a normal population. *Biometrika*, **32**, 301–10.

PEARSON, E. S. & ADYANTHAYA, N. K. (1928). The distribution of frequency constants in small samples from symmetrical populations. *Biometrika*, **20 A**, 356–60.

SIGNIFICANCE TESTS FOR 2×2 TABLES

BY G. A. BARNARD, *Imperial College*

PART I

The theory of statistical significance tests deals with abstractions of experimental results. The fact that the figures dealt with may happen to be tensile strengths of iron bars, or perhaps weights of babies, is ignored in the carrying out of the test; and for the purpose of statistical theory the experiment in question could just as well be represented by an experiment involving the drawing of balls from urns. In fact, it is an advantage, from some points of view, to replace the concrete experiment involved in a particular practical case by an 'abstract' urn-experiment, in order to retain in view only those features of the case which can be dealt with by statistical methods.

It is obvious enough that the first step in the statistical treatment of an experimental result may be represented as the replacement of the concrete experiment by an 'urn-experiment'; but the implications of this have not always had the continuous attention they deserve. Once the abstract picture has been formed, the analysis of it is largely a matter of pure mathematics. What distinguishes the statistician from the pure mathematician, in this connexion, should be the statistician's ability to form *valid* abstract pictures of concrete cases, and his clear recognition of the limits of validity of his abstract pictures. Yet we find relatively little discussion in statistical text-books of the process of formation of these abstract pictures.

It is the purpose of the first part of this paper to draw attention to the confusion which may arise through the possible formation of several different abstract pictures, each of which may apply to some concrete cases, though not to others.

Suppose we are given two mass-production processes, A and B, and we wish to test whether process A and process B are equally satisfactory, in the sense that neither process is more likely to produce defective items than the other. For this purpose we take, say, m articles made by process A, and n made by process B, and test them, under suitable conditions. We find that a out of the m articles are defective, while b out of the n articles are defective, a result which can be represented in the form of a 2×2 table (Table 1).

Table 1

	I (defective)	II (non-defective)	Total
Process A	a	c	m
Process B	b	d	n
Total	r	s	N

The statistical analysis of results of this type has been much discussed, but it seems to have escaped notice that, on the facts incompletely stated as above, it is possible to form several different abstract pictures, any one of which might be appropriate to the real case in question. The adoption of one picture rather than another will depend, in a given case, on further knowledge which is not specified above.

The basis of Fisher's 'exact' test

The current generally accepted test for results of the above type is that given by Fisher (1941), or some approximation to it. The simplest abstract picture * to which this test corresponds would seem to be one in which the m articles made by process A and the n articles made by process B are represented by N similar balls, m of them marked A and n marked B. The N balls are put into an urn, and then withdrawn in random order. As they are withdrawn, the balls are placed, in order, in a row of N receptacles, r of which have been marked 'I', the remainder being marked 'II'. The result of Table 1 then represents the observation that a of the balls marked A are in receptacles marked 'I'. The probability of such a result, in such an experiment is

$$\frac{m!\,n!\,r!\,s!}{N!\,a!\,b!\,c!\,d!} \tag{1}$$

which can be seen by considering that the contents of the r receptacles marked 'I' form a sample of r from an urn containing m balls marked A and n balls marked B, the sampling being done without replacement. The probability (1), added to those of all results less probable than that obtained, is the basis of Fisher's test.

In the concrete case given, the N balls, initially similar, may be taken to correspond with the N items of raw materials. The process of labelling the balls A and B corresponds to the selection of m of the items of raw material, and their fabrication into articles by process A, and the fabrication of the n remaining ones by process B. The N receptacles into which the balls are eventually placed then represent the N 'test occasions' which must be provided for when the experiment is laid out. The fact that these receptacles are labelled 'I' or 'II' before the balls are placed in them corresponds to the assumption of the hypothesis being tested—that the processes do not differ in respect of liability to defectives, so that whether or not a given article is defective has nothing to do with whether it is A or B. The labelling 'I' or 'II' is thus assumed independent of the labelling of the balls. Finally, the random allocation of balls to receptacles corresponds to a precaution which might have been taken in the concrete case, viz. the random order of test of the article secured by the use of random numbers or the like.

The basis of the C.S.M. test

Another abstract picture, also applicable to the concrete case as incompletely described above, forms the basis of the test to be developed in the later part of this paper, which we have called the C.S.M. test. In this picture, the two processes A and B, are represented by two urns, A and B, each urn containing a large number of balls, some of which are marked 'I', while the others are marked 'II'. The selection for test of m articles of process A is represented by the random drawing of m balls from urn A; and similarly for the n articles of process B. The test procedure corresponds to the examination of the balls, to see whether they are marked 'I' or 'II'. The liability of process A to produce defectives is represented by the proportion p_a of balls marked 'I' in urn A, while p_b similarly represents the liability of process B. The hypothesis we wish to test says that $p_a = p_b = p$, say. The probability of a result such as that of Table 1 is very nearly

$$\frac{m!}{a!\,c!}p_a^a(1-p_a)^c \times \frac{n!}{b!\,d!}p_b^b(1-p_b)^d \tag{2}$$

* Though not the only possible one. By following Fisher's argument, as given in his book, one can construct a more complicated picture which leads to a similar result.

which, on the hypothesis tested, becomes

$$\frac{m!\,n!}{a!\,b!\,c!\,d!}p^r(1-p)^s. \tag{3}$$

We may notice that the expression (3) differs from (1) by a factor

$$\frac{N!}{r!\,s!}p^r(1-p)^s$$

and it would have been obtained in the earlier case if we had assumed that the labelling of the receptacles was itself done randomly, by selection of N labels from a box containing a large number of labels, the proportion marked 'I' being p.

To justify the application of our second picture to a concrete case, we should have to be satisfied that the conditions of process A and those of process B were sufficiently stable, in a statistical sense, to justify the formation of the notions corresponding to p_a and p_b. We should further have to make sure that our selection of samples of m and n respectively was for practical purposes random. And finally, we should have to be reasonably sure that the conditions of test themselves had practically no influence on the results of the test—that the test used revealed a real property of the article tested, rather than a property of the individual conditions of test.

Another type of abstract experiment

Another case of common occurrence may be represented by a single urn, containing balls each of which carries two marks—one mark being either A or B, the other mark being either 'I' or 'II'. The experiment consists in drawing N balls from the urn, at random, and examining their markings. If the proportion of balls marked 'A I' is p_{a1}, while p_{b1}, p_{a2}, p_{b2} similarly represent the proportions of the other markings in the urn, the probability associated with Table 1 in this case is

$$\frac{N!}{a!\,b!\,c!\,d!}p_{a1}^a p_{b1}^b p_{a2}^c p_{b2}^d \tag{4}$$

by the multinomial theorem, provided the number of balls in the turn is large. In this case the hypothesis tested, that the markings 'I' and 'II' on the one hand, and the markings A and B on the other, are independent, may be put in the form

$$p_{a1}p_{b2} = p_{a2}p_{b1}$$

and, assuming that $(p_{a1}+p_{a2}) = p'$ and $(p_{b1}+p_{b2}) = 1-p'$, and $(p_{a1}+p_{b1}) = p$ and $(p_{a2}+p_{b2}) = 1-p$, do not vanish, the probability of our result, on the hypothesis tested, can be expressed as

$$\frac{N!}{a!\,b!\,c!\,d!}p^r(1-p)^s(p')^m(1-p')^n \tag{5}$$

which differs from (3) by a factor

$$\frac{N!}{m!\,n!}(p')^m(1-p')^n.$$

This shows that (5) is related to (3) in much the same way as (3) is related to (1).

This situation could present itself in our concrete case if the articles made by the two processes A and B were mixed up together in a common store, and the test sample of N were randomly drawn from this store, the subsequent conditions being as in the second case. Statisticians with industrial experience may perhaps feel it is unlikely that the experiment

would be performed in this way; but it must be admitted that it could have been. Cases such as this seem to occur more frequently in biometric investigations, where a population of animals is being tested for the association or otherwise of two characters.

Nomenclature

The name 'double dichotomy' has been applied generally to all experiments leading to results of the form of Table 1, but the foregoing analysis would suggest that it might be more appropriate to restrict this term to the third case we have indicated. Since the second case can be obtained from the third by supposing the numbers of articles made by process A and by process B to be fixed, we might then call the second case the (singly) restricted double dichotomy. Similarly, the first case would be called the doubly restricted double dichotomy. Such a nomenclature, apart from a lack of euphony, would be open to the objection that it would tend to imply that the third case was the general one, the first two being derivatives of it. This, in turn, would imply that the subject-matter of our investigation in cases one and two was in reality a four-fold universe, the restrictions on numbers being merely matters of experimental technique. But such is not always the case. The question implied in our second case presupposes two two-fold populations, which are to be compared, and no four-fold super-population need exist for this question to have meaning.

We therefore propose the names 'double dichotomy' for the third case, '2×2 comparative trial' for the second case, and '2×2 independence trial' for the first case, though here again an objection on aesthetic grounds would be easy to sustain.

Finer distinctions

In principle it could be maintained that there is a distinction between the 2×2 comparative trial, as instanced above, and a restricted double dichotomy. As we have said, the fundamental subject-matter of a 2×2 comparative trial is a pair of populations; while the subject-matter of a restricted double dichotomy is a four-fold population from which we happen, by an accident of experimental technique, to be able to extract samples in which the numbers of items having certain characteristics are fixed. The latter case could arise, for example, if an attempt was being made to discover association between colour of eyes in school-children and some less easily identified characteristic, such as membership of a particular blood-group. We could imagine that an experimenter might pick out m children with (say) blue eyes, and n without blue eyes, and then, having obtained his samples, he might subject them to a test for blood-group. The conclusions drawn from such an experiment would presumably be intended to apply to the population of school-children, a four-fold one relative to the two characteristics in question. The distinction between the two cases comes out if we consider what happened if, in the 2×2 comparative trial, all items tested turn out to be defective. In this case we should say that our question, whether $p_a = p_b$ or not, tends to be answered in the affirmative. In the case of the school-children, if they all turn out to have the same blood-group, then no conclusion on our question about the four-fold population can be drawn at all.

Similar distinctions apply to the 2×2 independence trial. In the psycho-physical experiment described by Fisher (1942), where the point at issue is whether or not a lady can tell whether the milk or the tea has been put in the cup first, no statistical population is presupposed. The question would have meaning even if we refused to regard the order of insertion of milk or tea as ever being a matter of chance, while at the same time we regarded

the lady's guess as equally determinate. The 'statistical population' enters into this experiment only in the experimental technique, via the randomization procedure used to fix the order of presentation of cups; it does not enter into the question being asked. In this case, the extreme result, in which in fact the milk was put in first every time, while the lady guessed every time that it was otherwise, would be taken as evidence against the lady's claim. But such a result could by itself have no meaning for the question asked in the case of a restricted 2×2 trial or a doubly restricted double dichotomy.

Further types of experimental procedure leading to results expressible in the form of Table 1 are the various sequential procedures that have been described for deciding questions of the kind we have been discussing (3, 4). Yet another procedure is one where the conditions of trial vary from one block of tests to another—as when an open-air trial runs over several days of inconstant weather. Here we might suppose there were k pairs of urns, (A_1, B_1), (A_2, B_2), ..., (A_k, B_k). The distinctions here are, however, obvious enough, and they are worth noting only in order to emphasize that the mere fact there results are presented in the form of Table 1 is not in itself sufficient to specify an appropriate test of significance.

PART II

The significance test for the 2×2 trial

Roughly speaking, the object of a significance test as applied to results of the type considered, is to answer the question: Can these results be ascribed to 'chance'? In this form, the question is not sufficiently precise. If our 'urn model' for the 2×2 comparative trial is adequate to represent the experiment actually carried out, then the results will in any case be 'due to chance', in some sense. What we wish to know in this case is whether a particular kind of chance—namely, one in which $p_a = p_b = p$—can be said to account for our results. If the results are such that this explanation of them is untenable, then we may conclude either, that our particular 'urn model' of the experiment is inadequate anyway; or we may retain the model, and conclude that p_a and p_b must be unequal. In most cases, of course, we shall reach the latter conclusion, since we would not have made up the urn model in question unless we had some reasons for believing in its adequacy; but it is well to bear in mind the first alternative, in case a re-examination of the circumstances may make us change our minds. A point very strongly emphasized by Fisher in his book *The Design of Experiments* is, that we ought to have in mind a particular 'urn model' *before* the experiment is performed, and arrange the conduct of the experiment so that the adequacy of this urn model is not likely to be questioned afterwards.

With the qualifications indicated, we can say that the object of the significance test we propose to develop is, to enable a particular class of explanations of our experimental results to be ruled out as untenable. Specifically, given results like those of Table 1, we want to be able to say that they could not be accounted for by supposing that the experiment we actually performed was analogous to the urn experiment with two urns in which $p_a = p_b = p$. This raises the question, in what sense could such a supposition fail to account for the observed results? Any result of the form of Table 1 *could* arise in an experiment of this kind, when our supposition is true. Why, then, should we select some results of this form and say they are incompatible with our supposition?

In the last analysis, this question cannot be answered without an examination of what is meant in general by statements involving probabilities, a point which is still the subject of

controversy. But in our particular case (if not in all cases) we can avoid giving a general answer to the question of what probability is, by considering the practical circumstances which form the setting for our particular problem, and the uses to which we propose to put the answer. In fact, in our case we are interested in the equality or otherwise of p_a and p_b because we want to decide which of the two processes, A and B, is to be preferred, from the point of view of defectives produced. To say that p_a is greater than p_b will mean, for us, that process B is preferable, and conversely if p_b is greater than p_a, while to say that p_a and p_b are equal will mean that there is nothing to choose between the two processes. In fact, to say that $p_a = p_b$, in our case, means that, if process A and process B are both used, then it will be found that the frequencies with which defectives appear in the two processes will, for practical purposes, be equal.* Thus we shall assert that results in which the observed frequencies, a/m and b/n, differ widely, are incompatible with the supposition that $p_a = p_b$; in doing so, we shall be neglecting as impossible a class of events which are in reality logically possible, but whose probability is small. The precise formulation of a test of significance then reduces to a precise formulation of what is meant by a 'wide difference' in the frequencies a/m and b/n, and to an evaluation of the probability of those events which are being neglected as impossible.

The lattice diagram

If we consider the first problem, of arranging results like those of Table 1 in order of the relative 'width' of the differences they indicate, a first step is the enumeration of all possible results in a convenient form.

Logically, we should begin by noting that Table 1 is really an abbreviated version of the results of any one particular experiment, which will to start with be like those of Table 2 (where we have taken $m = 8$, $n = 6$, for definiteness).

Table 2

Urn:	*A*	*A*	*A*	*A*	*A*	*A*	*A*	*A*	*B*	*B*	*B*	*B*	*B*	*B*
Mark:	II	I	II	II	I	II	II	II	I	I	II	I	I	I

But if, as we are presupposing, our urn analogy is adequate to represent the conditions of the experiment, the order in which the results were obtained must be irrelevant to the interpretation of results. If the conditions of trial varied during the course of the experiment, this assumption might not be correct—for example, if the trial were an open-air trial, and it began to rain half-way through. We are assuming that the urn analogy is adequate, and so we must treat all results like Table 2 which give the same values to a, b, c, d, in Table 1, as equivalent. Table 1 therefore stands for $m!\,n!/a!\,b!\,c!\,d!$ distinct, but equivalent, results which we shall not distinguish from now on.

If we now take rectangular axes in a plane, we can represent Table 1 by the point whose coordinates are (a, b). Thus 'x' in Fig. 1 represents the set of results equivalent, in the sense of the previous paragraph, to the results of Table 2. At the same time, all possible results of the experiment which gave rise to Table 2 are represented by the points of the rectangle

* We hope that the qualifications we have attached to our statements will be sufficient to guard us against the accusation that we have adopted in full a 'frequency theory' of probability. The frequency interpretation is relevant to our particular problem; other problems may involve other interpretations. More than one interpretation may be relevant in a single problem.

PQRS. We call this representation of possible results the lattice diagram.* Our problem may now be regarded as one of ordering the points of the lattice diagram according to the 'width' of the difference they indicate.

Conditions S and C

In trying to make the idea of 'width' of difference precise, we are up against difficulties similar to those attaching to the interpretation of results on the basis of incomplete information about the circumstances of the experiment. The information given at first was compatible with several distinct 'urn models'. Similarly, the information given now is compatible with several different notions about 'width' of difference. We may be concerned with the arithmetical size of the difference $p_a - p_b$, or with the ratio p_a/p_b, or with the logarithm of this ratio, or with some more complicated function.

Logically, therefore, we should expect to set up various tests, based on various ideas of what constitutes 'width' of difference in probability or frequency (in Neyman and Pearson's

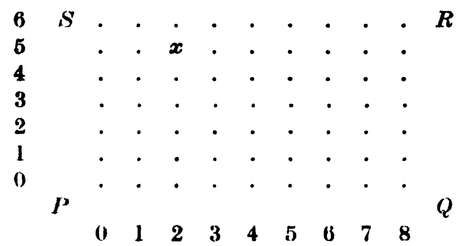


Fig. 1

Table 3

	I	II	Total
A	*c*	*a*	*m*
B	*d*	*b*	*n*
Total	*s*	*r*	*N*

language, corresponding to various weight functions over the space of alternatives to the hypothesis tested). But here a factor which may simply be described as laziness enters in. If we carried our ideas to their logical conclusion, we should find ourselves constructing a new test for almost every new experiment we had to deal with; and the time and effort involved in this are too great. Consequently, we confine our attempt to producing a test which will be reasonably applicable to a wide class of cases of the type specified, without suggesting that this test is unique, or 'best possible'.

First, then, in our ordering of points in the lattice diagram, we propose that the same rank should be given to the point $((m-a), (n-b))$ as to the point (a, b). This condition we propose to call the 'symmetry condition', or 'condition *S*'. It amounts to saying, that if Table 1 is to be considered as indicating a real difference between p_a and p_b, then so is Table 3, in which the labels 'I' and 'II' have been interchanged. If, when we are testing whether

* Not the sample space of Neyman and Pearson. In the sample space, different results equivalent to Table 2 are represented by different points.

$p_a = p_b$, we can say we are also testing whether $1-p_a = 1-p_b$, from the same point of view, then this symmetry condition is clearly justified.*

Next, we propose that in our ordering, the two points which, respectively, have the same abscissa or the same ordinate as (a, b), and which lie further from the diagonal PR, shall be considered as indicating wider differences than (a, b) itself. Thus, referring to Fig. 1, the points immediately above and immediately to the left of the point 'x' are reckoned to indicate wider differences than the point 'x' itself. This condition implies that the set of points indicating differences as wide or wider than (a, b) will have a shape property vaguely related to convexity, and we call it the 'C condition'. It means that if we consider the table corresponding to Table 2, with cell frequencies

$$\begin{matrix} 2 & 6 \\ 5 & 1 \end{matrix}$$

as significant evidence of difference, then we must also consider the tables

$$\begin{matrix} 1 & 7 \\ 5 & 1 \end{matrix} \quad \text{and} \quad \begin{matrix} 2 & 6 \\ 6 & 0 \end{matrix}$$

as significant evidence of difference. It is difficult to imagine circumstances where this would not be so.

Geometrically, condition S implies that we can in future restrict our considerations to points in the lattice diagram lying on or above the diagonal PR, i.e. in the triangle PRS. And condition C implies that, in this triangle, our 'width of difference' must increase as we go upwards or to the left. If horizontal and vertical axes are taken at any point X in this triangle, points in the second quadrant are associated with a wider difference than X is, points in the fourth quadrant are associated with narrower differences than X is. The relative width of differences associated with points in the first and third quadrants (excluding the axes) are now determined by the conditions C and S. The ordering generated by these conditions is thus a partial, not a total, ordering; it is, in fact, a kind of conical order, in the sense of A. A. Robb. We must introduce some further condition to make the ordering total.

Probability considerations

In many simpler cases, it is possible to distinguish those events which are considered incompatible with a given probability hypothesis by their relatively low probability, compared with other possible events. Such a simple comparison of probabilities is not open to us in this case, because to each point (a, b) we have, on the hypothesis tested, associated a function

$$W(a, b; p) = \frac{m!\,n!}{a!\,b!\,c!\,d!}\,p^r(1-p)^s$$

which contains the 'nuisance parameter' p. If we consider the relative position, in our ordering, of another point, (a', b'), we have to consider the inequality

$$W(a, b; p) < W(a', b'; p), \tag{6}$$

the truth or falsehood of which depends, in general, on the unknown p; and there is nothing in the statement of the problem, nor in the experimental method, to justify any particular choice for the value of p.

* Cases where $p_a > p_b$ is impossible are hereby neglected, strictly.

If $(a+b) = (a'+b')$, the validity or otherwise of the inequality (6) is independent of p. Thus, using this inequality as a criterion for ordering our points, we can say that in the triangle PRS, the 'width of difference' must increase as we move north-west. But this is all that can be derived from this criterion, and it is clearly even less helpful in ordering the points than the conditions C and S are. Moreover, if we recall that each point (a, b) in the lattice diagram really represents a set of $m!\,n!/a!\,b!\,c!\,d!$ distinct results, each with probability $p^r(1-p)^s$, the criterion (6) loses its plausibility.

We might try to improve the situation by associating the function $W(a, b; p)$ with a number, depending on a and b only. For fixed a and b, this number would be a functional of $W(a, b; p)$. We should clearly require that, if the inequality (6) is true for all p, then the corresponding inequality should be true of the numbers associated with $W(a, b; p)$ and $W(a', b'; p)$. The simplest functionals which satisfy this condition will be the mean value,

$$w(a,b) = \int_0^1 W(a,b;p)\,dp,$$

the maximum value $\quad w'(a,b) = \max\limits_{0<p<1} W(a,b;p),$

and one single value $\quad w''(a,b) = W(a,b;p_0).$

Circumstances could be imagined in which any of these three criteria might produce reasonable tests of significance. For example, in certain genetical experiments we may have reason to suppose that the value $p = 1/3$ would occur more often than any other value. In such a case we might use w'', with $p_0 = 1/3$. But for general purposes taking $p_0 = 1/3$ could not be justified.

We might again argue that taking w as our criterion would correspond with the assumption that all values of p were a priori equally likely. But some would say that such an assumption was never justified; while those who would admit the assumption would in strictness do so only if we really did know *nothing* about the value of p. And in the general circumstances we are trying to cater for, we may sometimes know something vague about the value of p—such as, for example, that p will be less than $\frac{1}{2}$.

Neyman and Pearson have shown that the likelihood ratio, which in our case comes to be

$$\frac{m^m n^n r^r s^s}{a^a b^b c^c d^d N^N}$$

very often gives a good basis for ordering experimental results. We feel, however, that the criterion we shall describe in the next section has a slightly more direct justification than the likelihood ratio, though the choice, is, admittedly, largely a matter of taste.

The maximum condition

Before setting out the final condition which, with conditions S and C, will be used eventually to arrange the points of the lattice diagram in order of 'relative width of difference indicated', we need to consider the assignment of significance levels to various results.

When we say that a given result is not significant on, say, the 5 % level, we mean that such a result, or one indicating a wider difference, could occur, with probability at least 0·05, even when $p_a = p_b$. We could believe in a theory that $p_a = p_b$, without having to suppose that an event belonging to a class whose joint probability was less than 0·05 had occurred. Conversely, if a result is judged significant on the 5 % level, it means that no theory which

assumed that $p_a = p_b$ could account for the result obtained without supposing that an event of a type whose probability was less than 0·05 had occurred on the occasion in question.

Let us now consider a specific case, in which we choose numbers which in practice would be ridiculously small in order to save arithmetic. Suppose, in fact, $m = n = 2$, while $a = 2$ and $b = 0$. It follows from conditions S and C alone that in judging the significance of such a result we need consider only the probability of this result, together with its converse, in which $a = 0$, $b = 2$. If $p_a = p_b = p$, the probability of results of this type is

$$P = 2p^2(1-p)^2.$$

Now suppose that we are prepared to discard as untenable theories which require us to suppose that events of probability less than 0·05 had occurred. In such a case, we should discard a theory which supposed $p_a = p_b = 0{\cdot}1$, since in this case $P = 2(0{\cdot}1)^2(0{\cdot}9)^2 = 0{\cdot}0162$, less than 0·05. But we could not discard a theory which supposed $p_a = p_b = 0{\cdot}5$, since in this case $P = 0{\cdot}125$. In fact, our result would enable us to discard all theories involving $p_a = p_b = p$, except those for which p lay in the interval $0{\cdot}197 < p < 0{\cdot}803$. In particular practical cases we might be prepared, on grounds external to the experiment in question, to dismiss the possibility that p should lie in this interval; and in such cases we should be entitled to say that the result excludes the possibility that $p_a = p_b$.

It is easy to see that the above specific case is typical. Any set of points in the lattice diagram, considered by some criterion agreeing with conditions S and C to indicate differences as wide or wider than those of a given result, will be associated with a probability P, on the assumption $p_a = p_b = p$; and this P will be a function of p, rising from zero when $p = 0$ to a maximum in the neighbourhood of $p = \frac{1}{2}$, and then falling again symmetrically (by the S condition) to zero again at $p = 1$, somewhat as in Fig. 2. The given result by itself

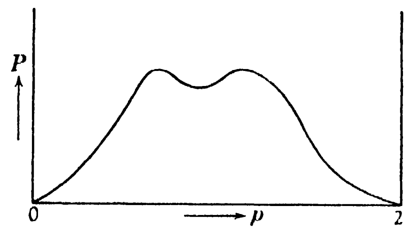


Fig. 2

will exclude the possibility $p_a = p_b$ altogether, only if the significance level adopted is greater than P_m, the maximum value of P. If our significance level corresponds to a probability less than P_m, then all we can say is, that our result is incompatible with $p_a = p_b$ unless their common value lies in a certain subset of the range (0, 1). We may or may not exclude these latter possibilities on other grounds.

In trying to construct our test, however, we have set ourselves the task of evaluating the evidence *provided by our experiment alone* in relation to the hypothesis $p_a = p_b$. It now appears that this is impossible so long as we restrict ourselves to the form, usual in such cases, of a simple statement that a given result is, or is not, significant on a given level. We have two alternatives. Either we can find an entirely new form of statement to convey what we wish to express; or we can adhere to the form of statement, and try to make the situation fit the form as nearly as possible. Perhaps the day will come when experimenters do not require answers in the form of numbers, when they are sufficiently versed in generalized mathematical analysis to be content with a function (such as the function $P(p)$), instead of a single

number. But we have not yet reached this stage; and so we propose to take up the latter alternative, and try to make the situation fit the standard form of statement of significance tests as nearly as possible.*

Our difficulty arises from the dependence of P on p. If the graph of P against p were a horizontal straight line, our difficulty would be overcome. What we propose, therefore, is to try to make the graph of P against p as near to a horizontal line as possible, by suitably adapting our idea of what is meant by 'width of difference'. In making this adaptation, we shall secure that we do not violate the common-sense requirements as to the meaning of the term 'width of difference', by requiring that conditions C and S should always be satisfied.

The maximum condition

The condition C requires that, of all points in the triangle PRS, that indicating the 'widest difference' must be the point S at the corner (Fig. 1). The function P associated with this point and its converse, Q, which we may denote as $P(0, 6; p)$, is

$$P(0, 6; p) = p^6(1-p)^8 + p^8(1-p)^6$$

and the maximum P_m occurs here when $p = \frac{1}{2}$, where we have

$$P_m(0, 6) = 1/2^{13} = 1{\cdot}22 \times 10^{-4}.$$

The condition C requires that the only points which might be considered as coming next after S, in order of decreasing 'width of difference' are $(1, 6)$ and $(0, 5)$. We have to adopt some principle to choose between these two.

If $(1, 6)$ were taken next after $(0, 6)$, the function P associated with it would be

$$P'(1, 6; p) = P(0, 6; p) + 16p^7(1-p)^7$$

and $P'_m(1, 6)$ would come to $9/2^{13} = 10{\cdot}97 \times 10^{-4}$. On the other hand, if $(0, 5)$ were chosen next, instead of $(1, 6)$, we should have

$$P(0, 5; p) = P(0, 6; p) + 6[p^9(1-p)^5 + p^5(1-p)^9]$$

and $P_m(0, 5)$ would come to $8{\cdot}58 \times 10^{-4}$, the maximum occurring when $p = \frac{1}{2} \pm \frac{1}{2}\sqrt{(6/70)}$. Thus $P_m(0, 5)$ is smaller than $P_m(1, 6)$, and this lower maximum is associated with a flatter curve of $P(0, 5; p)$. Since a flat curve is our aim (the horizontal line being the ideal), we choose $(0, 5)$ as the point to come next after $(0, 6)$, rather than $(1, 6)$.

Having chosen $(0, 5)$ as the next 'widest difference' point, the C condition restricts us to the points $(1, 6)$, and $(0, 4)$, as candidates for the next position. We consequently compare

$$P(1, 6; p) = P(0, 5; p) + 16p^7(1-p)^7$$

with
$$P''(0, 4; p) = P(0, 5; p) + 15[p^4(1-p)^{10} + p^{10}(1-p)^4]$$

and the lower value of P_m as criterion shows that $(1, 6)$ is now to be taken. At the next stage, we shall have to compare the functions associated with $(0, 4)$, $(1, 5)$ and $(2, 6)$. In this way we can arrange the points of the lattice diagram in order, step by step.

The principle involved, which we call the 'maximum condition', may be formally stated as follows:

Considering only points for which a/m is less than b/n, if the first $(n-1)$ points (a_1, b_1), (a_2, b_2), ..., (a_{n-1}, b_{n-1}), in order of decreasing 'width of difference' have been chosen, and

* In the example just taken we might make a kind of 'conditional confidence interval statement', that, if p existed, we should have $0{\cdot}197 < p < 0{\cdot}803$ with confidence coefficient $0{\cdot}95$.

(a_{n-1}, b_{n-1}) is associated with the function $P(a_{n-1}, b_{n-1}; p)$, then the nth point, (a_n, b_n) is that point, of all points (a, b) permitted by the C condition, for which

$$P_m(a, b) = \max_{0<p<1}\left[P(a_{n-1}, b_{n-1}; p) + \frac{m!\,n!}{a!\,b!\,c!\,d!}(p^r(1-p)^s + p^s(1-p)^r)\right]$$

is least. (a_n, b_n) is then associated with the function

$$P(a_n, b_n; p) = P(a_{n-1}, b_{n-1}; p) + \frac{m!\,n!}{a!\,b!\,c!\,d!}[p^r(1-p)^s + p^s(1-p)^r].$$

To complete the specification of the ordering, we have to legislate for the case where there are several points giving the same value of $P_m(a, b)$, this value being less than that associated with any other permissible point. In this case we lay down that all such points are to be given the same rank, and the second term in the expression for $P(a_n, b_n; p)$ is to be replaced by the corresponding sum over all these points. If there are k such points at any stage, then the next point after them will be denoted as the $(n+k)$th point in the ordering. This requires, for example, when $m = n$, that the points (a, b) and (b, a) are always to be taken together.

Finally, the significance level to be attached to the point (a_n, b_n) will be

$$P_m(a_n, b_n) = \max_{0<p<1} P(a_n, b_n; p).$$

This guarantees that our test will be a 'valid' one, in the sense that, if we judge a result incompatible with the hypothesis $p_a = p_b$, on a given level of significance, then *all* the possibilities of the form $p_a = p_b$ are excluded, to the given level. Thus no further information, external to the experiment in question, could make us decide that a result judged significant by our test was not in fact so (holding, of course, to a fixed significance level); on the other hand, we still have the possibility that other information may lead us to consider as significant results which appear in themselves not to be so. The formulation of our maximum condition is made so as to minimize this latter possibility. Our test is thus conservative, in the sense that we do not draw the conclusion $p_a \neq p_b$ unless this is certainly warranted by the data; but it might be called 'progressive conservative', because, of all such conservative tests, it will be the least conservative.

Another aspect of the maximum condition

When the author first approached the problem of analysis of experimental results of the type now considered, he did so from the point of view of regarding the significance level to be used as being fixed in advance, say at the 5 % level. From this point of view, the problem of constructing a test resolved itself, not into one of *ordering* the points in the lattice diagram, but into one of choosing a region, or set of points in the lattice diagram, such that any point belonging to this region could be regarded as evidence of inequality of p_a and p_b, on the given level of significance. The condition of symmetry required that such a region should consist of two similar parts, one above the diagonal PR, and one below it. The condition C required that the part of the region lying above the diagonal PR should be so shaped that if a point X belonged to the region, then so would all points lying north or west of X. There remained the problem, to decide which of the many regions satisfying these two conditions should be the one adopted.

To settle this, to any such region R we can associate a function

$$P(R; p) = \sum_{(a,b)\in R} \frac{m!\,n!}{a!\,b!\,c!\,d!} p^r(1-p)^s$$

and such a region will give a 'valid' test of significance provided that

$$\operatorname*{Max}_{0<p<1} P(R; p) \leqslant 0{\cdot}05.$$

There will not be so many regions satisfying this validity condition as well as the conditions S and C. We proposed, therefore, to select that region from among these, which had the greatest number of points in it. This last condition was what we then called the 'maximum condition'. The fact that this region would not be unique in cases where $m = n$ was taken care of by requiring a subsidiary symmetry condition that in such cases (a, b) and (b, a) should always be taken together.

What we have now adopted as the 'maximum condition' can be seen to be related to this earlier version, by the consideration that, roughly speaking, apart from effects due to the discreteness of the lattice diagram, holding the number of points in the region constant, and then choosing the region which gives the lowest value for P_m, as we do now, comes to the same thing as holding P_m constant, and then choosing the region to have the maximum number of points.

Other things being equal, the 'power' of a test, in the sense of Neyman and Pearson, will increase with the 'volume' of the rejection region chosen. In this sense we can say, roughly, that the maximum condition secures that our test should be as powerful as possible, consistent with validity.

Practical formulation of the test

Some statistical tests (such as that due to Fisher, already mentioned), can be carried out in the form of a direct calculation from the data, without reference to any special tables. Most other tests require the use of special tables which, however, are for the most part tables of single or double entry, perhaps triple entry, if the level of significance is regarded as a variable. In our case, regarding the level of significance as a variable, a table of quadruple entry would be required.

Ideally, a set of tables, one for each pair of values of m and n $(m \geqslant n)$ would be required. The table would be in the shape of a right-angled triangle, corresponding to the triangle PRS of Fig. 1, and divided into squares, each square corresponding to given values of a and b. Within each square (a, b) would then appear a number, the value of $P_m(a, b)$. This value of $P_m(a, b)$ then is the maximum probability of obtaining the result (a, b), or one indicating a wider difference, if $p_a = p_b$. A comparison of $P_m(a, b)$ with the significance level adopted will then decide the significance or otherwise of our result. In any particular case we shall be able to see which tables, in the sense of our test, are regarded as indicating a wider difference, by noting which points are associated with lower values of $P_m(a, b)$.

In practice, it will be impossible to construct such tables for a large range of values of m and n. But for larger values of m and n, a test based on a normal approximation to the distributions involved will be quite adequate for practical purposes. In fact, the test we have proposed will itself approximate, in some sense, to a test based on the normal distribution, though we do not enter into a detailed discussion of the relationship between the two tests here.* Tables are thus required for our test only for small values of m and n. In spite of advice by statisticians to the contrary, such small values of m and n continue to occur frequently in practice.

* The general question of the sense in which tests are regarded as 'asymptotically approaching' normal tests is a subject for another paper. Professor Pearson's paper which follows, bears on this point.

In the Appendix we give specimen tables for the cases where $N = 14$. The comparative figures for the Fisher test, also given in the Appendix, indicate that the differences between the two tests are appreciable. An exploration is now under way into larger values of m and n, and it is hoped to report on this in due course.

Other applications of the C.S.M. procedure

We have spoken of our test as *the* C.S.M. test, as if the case dealt with above were the only case to which the procedure adopted was applicable. But similar methods could be used in many other cases. In particular, a method closely following the one we have used might be applied to the case we have called the double dichotomy, which differs from the 2×2 comparative trial in that two 'nuisance parameters', p and p' are present, instead of only one. The 2-dimensional lattice diagram of the 2×2 trial is replaced by a 3-dimensional regular tetrahedron of points with homogeneous coordinates (a, b, c, d), connected by the relation

$$a + b + c + d = N.$$

Two opposite edges of this tetrahedron correspond to $m = 0$ and $n = 0$, and sections of the tetrahedron by planes parallel to these edges will look exactly like lattice diagrams for the 2×2 case and within these sections, relative probabilities will behave just as in the 2×2 case. An examination of the possibilities, however, indicates that not much is to be gained by a detailed treatment. The C.S.M. test for 2×2 comparative trials will be a valid test if applied to double dichotomies. It will err somewhat on the side of 'conservatism', but the error does not appear to be large, except when the numbers involved are exceedingly small.

It is with a view to further applications of the approach used in this paper that we have retained the C condition as a separate requirement, although it is easy to see that it could be absorbed into the M condition as we have given it.

In writing this paper the author has had great personal help and encouragement from Prof. E. S. Pearson, to whom he wishes to express his very deep thanks.

Summary

In Part I we discuss various types of experiment, each of which may give rise to results in the form of a 2×2 table. It appears that significance tests which may be appropriate for one type of experiment will not necessarily be appropriate for another.

In Part II a test is developed for experiments of the type called '2×2 comparative trials'.

APPENDIX

Tables for the CSM test

Three tables are given below to illustrate the application of the ideas given in the main paper to the construction of a test for 2×2 comparative trials. The cases covered are pairs of samples, sizes (7, 7), (8, 6), and (9, 5). The small figures in brackets in the (7, 7) table gives significance levels on Fisher's 'exact' test for 2×2 independence trials, for comparison. Only half of the (8, 6) and (9, 5) tables are given; the missing parts can be filled in by symmetry. The following examples show the meaning and use of the tables:

Example 1. Two boxes, each containing a large number of components, are to be tested for comparative quality measured by the respective proportions of defective components they contain. Two samples, each of seven components, are taken, at random, one from each box. One sample gives four defectives, the other, none. What is the significance of this result, in relation to the hypothesis that the boxes have the same quality?

Answer. Entering the (7, 7) table at the point (0, 4), we find the number 2·4. This means that the result is evidence against the hypothesis, on the 2·4 % level of significance.

Table for $m = n = 7$

7	0·012 (0·058)	0·18 (0·23)	0·70 (2·1)	2·4 (7·0)	7·5 (19)	20 (46)	—	—
6	0·18 (0·23)	1·3 (2·9)	5·7 (10)	13 (27)	—	—	—	—
5	0·70 (2·1)	5·7 (10)	21 (29)	—	—	—	—	20 (46)
4	2·4 (7·0)	13 (27)	—	—	—	—	—	7·5 (19)
3	7·5 (19)	—	—	—	—	—	13 (27)	2·4 (7·0)
2	20 (46)	—	—	—	—	21 (29)	5·7 (10)	0·70 (2·1)
1	—	—	—	—	13 (27)	5·7 (10)	1·3 (2·9)	0·18 (0·23)
0	—	—	20 (46)	7·5 (19)	2·4 (7·0)	0·70 (2·1)	0·18 (0·23)	0·012 (0·058)
	0	1	2	3	4	5	6	7

More precisely, what is asserted is, that the maximum probability of getting a result not less significant than that obtained, is 0·024. And the results which are not less significant are those which correspond to points in the table with numbers not greater than 2·4, viz. (0, 4), (7, 3), (0, 5), (7, 2), (0, 6), (7, 1), (0, 7), (7, 0), (1, 6), (6, 1), (1, 7), (6, 0), (2, 7), (5, 0), (3, 7), (4, 0). By suitable choice of the proportion defective, we could construct a pair of boxes, of equal quality, which would give samples falling in this group 24 times out of 1000, on the average; but we could not, by any choice of proportion defective, retain equal quality and yet have results in this group more often than 24 times in 1000.

Table for $m = 8$, $n = 6$

	0	1	2	3	4	5	6	7	8		
6	0·012	0·18	0·71	2·5	5·3	13	—	—	—	—	
5	0·085	1·3	6·6	11	—	—	—	—	—	—	5
4	0·44	3·9	19	—	—	—	—	—	—	—	4
3	1·9	16		—	—	—		—	—	6·3	3
2	8·0	—	—	—	—	—	—	20	3·8	0·86	2
1	23	—	—	—	—		14	7·4	1·3	0·13	1
0	—	—	—	16	10	5·3	2·3	0·62	0·19	0·012	0
	0	1	2	3	4	5	6	7	8	9	

Table for $m = 9$, $n = 5$

Example 2. The situation is as before, except that the first sample has nine components, none of them defective, while the second sample has five components, four of them defective.

Answer. Here, to use the table as given, we have to compare numbers effective, rather than numbers defective—viz. we consider the pair (9, 1) rather than (0, 4). Entering the (9, 5) table at (9, 1) we find 0·13. The result is evidence against the hypothesis of equal quality, on the 0·13 % level of significance.

Thanks are due to Miss Lang, who has checked the computations.

REFERENCES

FISHER, R. A. (1941). *Statistical Methods for Research Workers*, 8th ed. Edinburgh: Oliver and Boyd.
FISHER, R. A. (1942). *The Design of Experiments*, ch. II. Edinburgh: Oliver and Boyd.

THE CHOICE OF STATISTICAL TESTS ILLUSTRATED ON THE INTERPRETATION OF DATA CLASSED IN A 2 × 2 TABLE

By E. S. PEARSON

CONTENTS

(i) Introductory

1. The problem of testing the significance of a difference between two proportions is one which receives early attention in text-books on mathematical statistics, and it might be thought to be one of the questions whose final solution lies behind us. It is a problem whose simplicity makes it easy to examine the logical cogency of the methods put forward for its solution, but, on examination, it is evident that they have not yet been rounded off satisfactorily. The origin of the present paper lies partly in an investigation commenced in 1938 and discussed at the time in College lectures, and partly in recent correspondence in *Nature* in which G. A. Barnard (1945*a*, *b*) and R. A. Fisher (1945*a*) have taken part.* This correspondence has suggested that in a problem of such apparent simplicity, starting from different premises, it is possible to reach what may sometimes be very different numerical probability figures by which to judge significance.

2. Such a difference in levels of significance in the solution of an everyday problem is obviously puzzling to the users of statistical methods who are accustomed to accept the technique as an established procedure and have not the opportunity for a critical examination of the conditions under which probability theory is brought to bear as a guide to action. For the question here at issue is a fundamental one of why and how our judgement is influenced by the calculation of a probability, and the dilemma raised by the Barnard-Fisher correspondence can only be answered in terms of our views on the practical function of the theory. We may all agree that in practice we use probability figures derived from an analysis of numerical data to help us to make up our minds on the next step, whether in experimental research or executive action. But what form of presentation of the probability set-up is likely to result in the greater number of sound decisions is likely to be always a matter for differences of opinion.

3. All that I can do is to approach the problem of the 2 × 2 table from the viewpoint which appears most helpful to me. In the preceding paper Mr Barnard has elaborated the

* There was also an earlier discussion on the same subject between E. B. Wilson (1941, 1942) and R. A. Fisher (1941).

views expressed in his letters to *Nature*. Such discussion is, I believe, desirable, even though controversial issues are raised. For the value of the whole elaborate structure of the modern theory of mathematical statistics depends at least in part on the sense in which the individual statistician appreciates the meaning of the probability model he is using when drawing the practical conclusions from his analysis of data. I have used the words 'in part', for it is true that the analytical process of applying the statistical technique to experimental data may in itself be enormously illuminating even without paying any close regard to a final probability figure. Such is the case, for example, with the technique of analysis of variance, where the mere process of breaking up a total sum of squares into parts with which different sources of variability can be associated, brings with it a reward in clear thinking even without the application of a probability test.

4. There is a very wide variety in the types of situation in which probability theory is introduced to help in reaching a decision as to further action.

(A) At one extreme we have the case where repeated decisions must be made on results obtained from some routine procedure carried out under controlled conditions.

(B) At the other is the situation where statistical tools are applied to an isolated investigation of considerable importance in which many of the issues involved in the conclusion can hardly be assessed in numerical terms.

5. Two situations of this kind, in which the statistical technique involved is that of testing the significance of a difference between two proportions, may be illustrated from problems arising in the 'proof' of armour-piercing shot or shell.

6. *Example of type A*. In the proof of small anti-tank, armour-piercing shot it might be decided to set aside, as a standard, a batch of shot whose quality has been established by special trials; against this standard, later batches can be compared. The variable measured is the proportion of shot which fail to perforate a plate of specified thickness when fired with a given striking velocity. The use of standard shot is necessary for calibration purposes, because there are inevitable changes in toughness from one proof plate to another and only a limited number of shot can be fired at a single plate. Then the situation might be summed up as follows:*

Aim of proof. To ensure that as few batches as possible are passed into service which are less effective than the standard.

Method of proof. Twelve rounds of the standard and twelve of the batch under test to be fired, round for round, against a single test plate and a record kept of the number of failures in each group, say a and b.

Routine sentencing rule. This should lay down a ready means of determining, from a knowledge of a and b, whether to class the new batch as inferior to the standard or not.

Assumptions accepted in using rule. That the two samples of twelve shot have each been randomly selected from the much larger batches. That against the particular plate used, a proportion p_1 of the standard and p_2 of the new batch would fail to give satisfactory perforation at the specified striking velocity. That while p_1 and p_2 would be different for other plates, if $p_2 > p_1$ for one plate, it will be so for all other plates. The objective is to segregate batches of shot for which $p_2 > p_1$.

* It has been somewhat simplified for illustrative purposes, e.g. complete control of the striking velocity is not in practice possible.

7. *Example of type B.* Two types of heavy armour-piercing naval shell of the same calibre are under consideration; they may be of different design or made by different firms. Since the cost of producing and testing a single round of this kind runs into many hundreds of pounds, the investigation is a costly one, yet the issues involved are far reaching. Twelve shells of one kind and eight of the other have been fired; two of the former and five of the latter failed to perforate the plate. In what way can a statistical test contribute to the decision which must be taken on further action?

8. In dealing with Example A the guiding principle followed in seeking help from the theory of probability can be very simple. We can set as our object a rule which:

(i) will result in an increasing chance of detecting that $p_2 > p_1$, the larger the difference;

(ii) will leave only a small chance of segregating the new batch wrongly when, in fact, $p_2 \leq p_1$.

Diagrammatically the rule would consist in segregating the new batch when the point (a, b) falls within some such area as that shown shaded in Fig. 1. In this problem involving a routine procedure, it is the long-run frequency of different consequences of the proof sentencing which is of importance, and probability theory is introduced to provide a measure of expected frequency. This method of introducing the theory of probability into this proof problem is not necessarily the only one that could be adopted in fixing a routine procedure, but it is a simple one and, since simplicity has the merit of appealing to the user's understanding, it has great advantages.

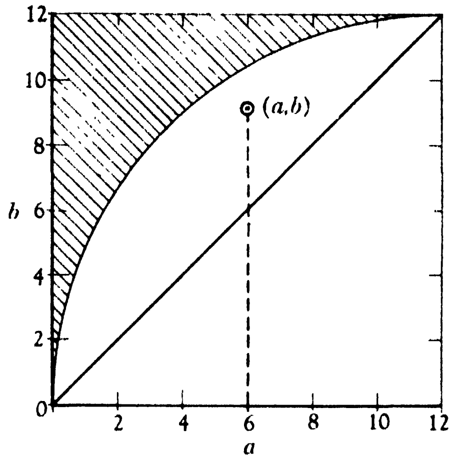


Fig. 1

9. When dealing with Example B a very considerable number of factors must be weighed in the balance, and the result of a statistical test of significance could never be the over-riding one. There will be other information as to the effect of changes in shell design, possibly from shell of different calibre; information as to the uniformity in quality of output of the firm or firms concerned; questions of cost and of general policy. He would be a bold man who would attempt to express these in numerical terms. Whereas when tackling problem A it is easy to convince the practical man of the value of a probability construct related to frequency of occurrence, in problem B the argument that 'if we were to repeatedly do so and so, such and such result would follow in the long run' is at once met by the common-sense answer that we never should carry out a precisely similar trial again.

10. Nevertheless, it is clear that the scientist with a knowledge of statistical method behind him can make his contribution to a round-table discussion, provided he has acquired a grasp of the practical issues. Starting from the basis that individual shell will never be identical in armour-piercing qualities, however good the control of production, he has to consider how much of the difference between (i) two failures out of twelve and (ii) five failures out of eight is likely to be due to this inevitable variability. There may be a number of ways of sizing up the position involving different assumptions or hypothetical constructs; he may follow one or several of these. The value of his advice is dependent almost

entirely on the soundness of his scientific judgement, and very little on whether his backroom calculations have been based on inverse or direct probability or on an appeal to fiducial argument.

11. How far, then, can one go in giving precision to a philosophy of statistical inference? It seems clear that in certain problems probability theory is of value because of its close relation to frequency of occurrence; such seems to be the case for my Example A. Tests can be built up to satisfy the practical requirements in this field. In other and, no doubt, more numerous cases there is no repetition of the same type of trial or experiment, but all the same we can and many of us do use the same test rules to guide our decision, following the analysis of an isolated set of numerical data. Why do we do this? What are the springs of decision? Is it because the formulation of the case in terms of hypothetical repetition helps to that clarity of view needed for sound judgement? Or is it because we are content that the application of a rule, now in this investigation, now in that, should result in a long-run frequency of errors in judgement which we control at a low figure? On this I should not care to dogmatize, realizing how difficult it is to analyse the reasons governing even one's own personal decisions.

12. That the frequency concept is not generally accepted in the interpretation of statistical tests is of course well known. With his characteristic forcefulness R. A. Fisher (1945*b*) has recently written: 'In recent times one often repeated exposition of the tests of significance, by J. Neyman, a writer not closely associated with the development of these tests, seems liable to lead mathematical readers astray, through laying down axiomatically, what is not agreed or generally true, that the level of significance must be equal to the frequency with which the hypothesis is rejected in repeated sampling of any fixed population allowed by hypothesis. This intrusive axiom, which is foreign to the reasoning on which the tests of significance were in fact based seems to be a real bar to progress....'

13. But the subject of criticism seems to me less an intrusive mathematical axiom than a mathematical formulation of a practical requirement which statisticians of many schools of thought have deliberately advanced. Prof. Fisher's contributions to the development of tests of significance have been outstanding, but such tests, if under another name, were discovered before his day and are being derived far and wide to meet new needs. To claim what seems to amount to patent rights over their interpretation can hardly be his serious intention. Many of us, as statisticians, fall into the all too easy habit of making authoritative statements as to how probability theory should be used as a guide to judgement, but ultimately it is likely that the method of application which finds greatest favour will be that which through its simplicity and directness appeals most to the common scientific user's understanding. Hitherto the user has been accustomed to accept the function of probability theory laid down by the mathematicians; but it would be good if he could take a larger share in formulating himself what are the practical requirements that the theory should satisfy in application.

(ii) THE CHOICE OF STATISTICAL TESTS

14. One approach to follow in determining tests to be applied to the 2×2 class of problem follows the lines that Neyman and I have adopted since 1928 in dealing with tests of statistical hypotheses. Let me first recapitulate in broad terms the steps in that approach when applied to a problem where the universe of possible observations can be represented by a

finite set of discrete points. A test of significance may be described as a method of analysis of statistical data which helps us to discriminate between alternative theories or hypotheses. In order to make use of the theory of probability in the sense here understood, a random process must either have been purposely introduced or be assumed to have been present in the collection of data; then the hypothesis very often concerns the values of parameters contained in the probability laws which, in the conceptual sphere, form the mathematical counterpart of the sampling distributions of experience.

15. We proceed by setting up a specific hypothesis to test, H_0 in Neyman's and my terminology, the null hypothesis in R. A. Fisher's. At the same time, in choosing the test, we take into account alternatives to H_0 which we believe possible or at any rate consider it most important to be on the look out for. Thus we wish the test to have maximum discriminating power within a certain class of hypotheses. Three steps in constructing the test may be defined:

Step 1. We must first specify the set of results which could follow on repeated application of the random process used in the collection of the data; this may be termed the experimental probability set.

Step 2. We then divide this set by a system of ordered boundaries or contours such that as we pass across one boundary and proceed to the next, we come to a class of results which makes us more and more inclined, on the information available, to reject the hypothesis tested in favour of alternatives which differ from it by increasing amounts.

Step 3. We then, if possible, associate with each contour level the chance that, if H_0 is true, a result will occur in random sampling lying beyond that level.

This rather crude statement of procedure will be developed in more detail in discussing the problems that arise in connexion with the 2×2 table.

16. *Notes on these points.* (*a*) *Step* 1. This involves the definition of what Neyman and I have termed the sample space, W. The application in three forms of the 2×2 problem is discussed in paragraphs 19, 27 and 46 below.

(*b*) *Step* 2. For a given hypothesis under test there may be a number of ways of deriving a system of contours, and only in certain cases can there be said to be complete agreement on which is the 'best'. Practical expediency will often carry weight in the choice. It is widely accepted that the choice cannot be made without paying regard to the admissible hypotheses alternative to H_0, whether this process is given formal precision or taken as a broad guide. In our first papers (Neyman & Pearson, 1928*a*, *b*) we suggested that the likelihood ratio criterion, λ, was a very useful one to employ in determining a family of contours which would be ordered in relation to our confidence in the hypothesis tested when set against the background of admissible alternatives. Thus Step 2 preceded Step 3. In later papers (Neyman & Pearson, 1933, 1936 and 1938) we started with a fixed value for the chance, ϵ, of Step 3 and determined the associated contour, taking account of what we termed the power of a test with regard to the alternative hypotheses. The family of Step 2 followed on giving decreasing values to ϵ. However, although the mathematical procedure may put Step 3 before 2, we cannot put this into operation before we have decided, under Step 2, on the guiding principle to be used in choosing the contour system. That is why I have numbered the steps in this order.

(*c*) *Step* 3. If this can be accomplished, we have what Neyman and I called control of the '1st kind of error'. In problems where, as below, we are concerned with discrete rather than

continuous probability distributions (e.g. for the binomial, the Poisson, the multinomial and the hypergeometric distributions), this objective cannot always be achieved, and it may be necessary to be satisfied with a knowledge of an upper limit of the chance of rejecting the hypothesis tested when it is true.

(iii) APPLICATION OF THIS APPROACH TO THE ANALYSIS OF DATA CLASSED IN A 2×2 TABLE

17. The frequencies of the data in the table may be defined in the following notation:

Table 1

	Col. 1	Col. 2	Total
Row 1	a	c	m
Row 2	b	d	n
Total	r	s	N

If we follow in turn the steps defined above to determine the method of interpretation of such data, the requirements of the appropriate tests are seen to follow very simply, although mathematical or computational difficulties arise in implementing them. On taking Step 1 we can separate out at once the three types of problem which Barnard has differentiated;* these I shall call Problems I, II and III. They are distinguished by the sample space having 1, 2 and 3 dimensions respectively. From the mathematical point of view it might seem more logical to take them in the reverse order, adding first one and then a second restriction to the 3-dimensioned case of Problem III. For a simple exposition, I think the reverse procedure of building up from I to III is preferable and this has been adopted in the following sections.

(iv) PROBLEM I

18. This may be described as the test of the significance of the difference between two treatments after these have been randomly assigned to a group of $N = m+n$ individuals (Barnard terms it the 2×2 independence trial). To use the terminology of a particular application, we may say that we are observing the presence or absence of 'reaction X'. The first treatment is applied to m and the second to n of the N individuals; as a result a/m and b/n show reaction X.

19. In this case the random process has been applied within the group of N individuals, and its repetition would simply involve other random reassignments of the two treatments among the N. No assumption is made as to how the N individuals were selected from some larger universe. The repetition may be hypothetical, in the sense that it often could not take place, e.g. if reaction X = death. Indeed, repetition under the same essential conditions is frequently impossible in practice. But this correspondence between the frequency of results upon hypothetical repetition and the probability distribution of the counterpart mathematical model forms an accepted part of the process of reasoning whereby (following

* Statisticians had, of course, all been more or less conscious of these differences, but, at any rate in my own case, it was discussion with Mr Barnard which made it easy to see the problem in its full clarity.

the present approach) we use probability theory as a basis for inference. The hypothesis tested is that while some individuals show reaction X and some do not, the result would be the same whichever treatment were applied *as far as these N individuals are concerned.* Thus, on the null hypothesis, there are $r = a+b$ individuals who will react and $s = c+d$ who will not, whatever the assignment of treatments.

20. The chance that a will react in m and $b = r-a$ in n is, therefore, if the hypothesis be true,

$$P_1\{a \mid N, r, m\} = \frac{m!\,n!\,r!\,s!}{a!\,b!\,c!\,d!\,N!}. \tag{1}$$

This expression is proportional to the coefficient of x^a in the hypergeometric series

$$F(\alpha, \beta, \gamma, x) = F(-r, -m, n-r+1, x). \tag{2}$$

Thus, taking $m \geqslant n$, a can assume values of

$$\text{(i)}\quad 0, 1, \ldots, r \quad \text{if} \quad r \leqslant n,$$

$$\text{(ii)}\quad r-n, r-n+1, \ldots, r \quad \text{if} \quad n < r \leqslant m,$$

$$\text{(iii)}\quad r-n, r-n+1, \ldots, m \quad \text{if} \quad r > m.$$

For this probability distribution, it is known (K. Pearson (1899) and Kendall (1943, p. 127))

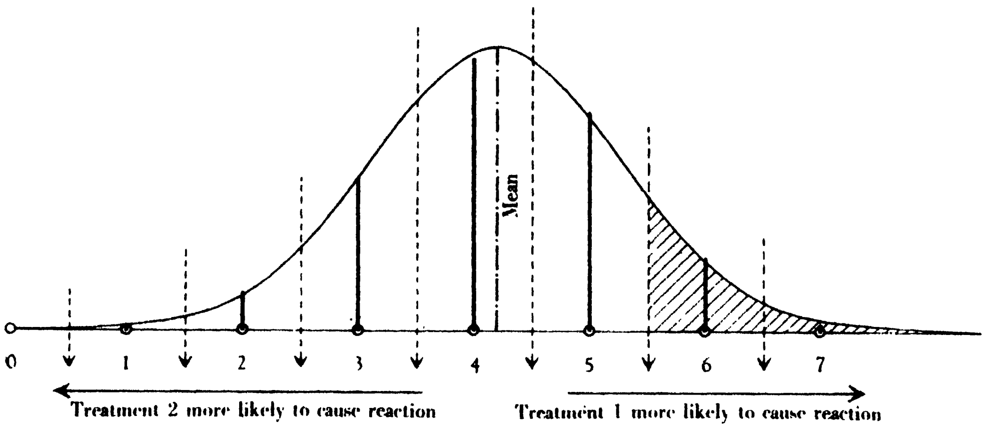


Fig. 2

that

$$\text{Mean}\, a = \frac{rm}{N}, \tag{3}$$

$$\text{Variance of } a = \sigma_a^2 = \frac{mnrs}{N^2(N-1)}. \tag{4}$$

21. For the particular case

$$N = 20, \quad r = 7, \quad m = 12, \quad n = 8,$$

the terms in the distribution of $P_1\{a \mid 20, 7, 12\}$ are shown as ordinates in Fig. 2 and given in the accompanying Table 2. The experimental probability set consists of the eight alternative values for a, viz. $0, 1, \ldots, 7$ with which the probabilities tabled are associated if H_0 is true. Further

$$\text{Mean}\, a = \bar{a} = 4{\cdot}2, \quad \sigma_a = 1{\cdot}0721. \tag{5}$$

22. Next consider step 2. The purpose of the investigation is to test the hypothesis that the difference between $a/12$ and $(r-a)/8$ has resulted simply from a random partition of 20 individuals, of whom r will show reaction X in whichever treatment group they are included. The experiment gives $r=7$. The contour levels fall between the 8 points of the set as shown in Fig. 2; the further a lies towards the right, the more inclined we shall be to accept the alternative hypothesis that $a/12>(r-a)/8$ because treatment 1 is more effective than treatment 2. The further a lies to the left, the more we shall incline towards the reverse alternative. To complete Step 3, we have only to calculate the sums of the tail terms of the hypergeometric series, as shown in Table 2 for the special case.

Table 2. *Problem I. Chances for special case* $N = 20$, $r = 7$, $m = 12$, *if* H_0 *is true*

a	Chance of a	Chance of a or less	
		True value	Normal approx.
0	0·0001	0·000	0·000
1	0·0043	0·004	0·006
2	0·0477	0·052	0·056
3	0·1987	0·251	0·257
4	0·3576	—	—
		Chance of a or more	
		True value	Normal approx.
5	0·2861	0·392	0·390
6	0·0954	0·106	0·113
7	0·0102	0·010	0·016

23. Having set up the machinery of the test, we come to the practical question. Beyond which contour levels must a fall before we infer that there is a treatment difference? Not, I think, in the example, if a were 3, 4 or 5; possibly if $a = 6$, more probably if $a = 2$ and almost certainly if $a = 0$, 1 or 7. Were we to fix as critical levels those between $a = 1$ and 2 on the one hand, and between $a = 6$ and 7 on the other, then we should be guided in our decision by the following knowledge: if there were no treatment difference, so that seven out of the twenty individuals would have shown reaction X whichever treatment were applied, then the chance under random assignment of treatments that $a<2$ or >6 is only 0·014 or 1 in 70. Had we taken the critical levels between 2 and 3 and between 6 and 7, the corresponding chance would be 0·062 or 1 in 16. This summing up in terms of probability helps towards the balanced decision on the next practical step to be taken, because it helps us to assess the extent of purely chance fluctuations that are possible. It may be assumed that in a matter of importance we should never be content with a single experiment applied to twenty individuals; but the result of applying the statistical test with its answer in terms of the chance of a mistaken conclusion if a certain rule of inference were followed, will help to determine

the lines of further experimental work and the degree of confidence with which we proceed provisionally to adopt a new technique.

24. An experiment falling under this head has the advantage that the random process introduced is under complete control. The analysis will give an answer in probability terms whether the N individuals have been randomly selected from a larger whole or not. But this answer is limited in the sense that it relates only to the N; if we wish to draw conclusions about a wider population or populations, then a random selection of the N or, separately, of both its parts m and n is needed. Thus we come to Problems II and III.

25. *Approximation to the hypergeometric terms.* When dealing with small numbers, the calculation of the tail terms of the series may not be laborious, but it soon becomes so when r is large. An obvious approximation is that obtained by using an integral under the normal curve with the mean and standard deviation of equations (3) and (4) to represent the sum of the hypergeometric terms. As usual when approximating to the sum of the terms for $x = a, a+1, a+2, \ldots$, etc., of a discrete probability distribution by the integral under a continuous curve, we take this integral from the point $x = a - \frac{1}{2}$. Thus Fig. 3 shows the normal curve

$$p(x) = \frac{1}{\sqrt{(2\pi)}\,\sigma_a} \exp\left[-\tfrac{1}{2}(x-\bar{a})^2/\sigma_a^2\right], \tag{6}$$

with $\bar{a}$ and σ_a as in equations (5), and the approximation to the sum of the hypergeometric terms for $a = 6$ and 7 is

$$\int_{5.5}^{\infty} p(x)\,dx,$$

represented by the area marked with cross-hatching. The approximations for different levels are shown in Table 2, and are seen in this case to be quite adequate for the purpose of the test. Further comparisons are made in the Appendix, and it appears that provided m and n are fairly nearly equal, as they are likely to be in most planned experiments of the Problem I type, the normal approximation is surprisingly good. Yates (1934) has suggested a method of further correction.

26. *The correction for continuity.* In the 2×2 table connexion, the improvement obtained by taking the normal integral (i) from $x = a - \frac{1}{2}$ if $a > \bar{a}$ or (ii) from $x = a + \frac{1}{2}$ if $a < \bar{a}$ (so that we are summing for the lower tail), was pointed out by Yates (1934) and has often been termed 'Yates's correction for continuity'. It is, however, the natural adjustment to make on the basis of the Euler-Maclaurin theorem, when approximating to a sum of ordinates by an integral and without wishing to detract from the value of Yates's suggestion in this particular problem, it should be pointed out that the adjustment was used by statisticians well before 1934, when employing a normal or skew curve to give the sum of terms of a binomial or hypergeometric series.*

(v) PROBLEM II

27. This may be described as the test of whether the proportion of individuals bearing a character A is the same in two different populations, from each of which a random sample has been drawn, i.e. the test of the hypothesis that

$$p_1(A) = p_2(A) = p, \tag{7}$$

* The method was in use in the Department of Applied Statistics when I joined the staff in 1921, and may have been current many years before that.

where p is some common but unspecified proportion. Barnard describes this as the case of the 2×2 comparative trial. Here m individuals have been drawn at random from the first population and n from the second, and it is found that a/m and b/n, respectively, bear the character A. The conditions are assumed to be such that if the random procedure of selection were repeated, the appropriate probability distributions for a and b would be given by the terms of binomial expansions. Table 3 shows the observed results.

Table 3

	No. with character A	No. without A	Total
1st sample	a	c	m
2nd sample	b	d	n
Total	r	s	N .

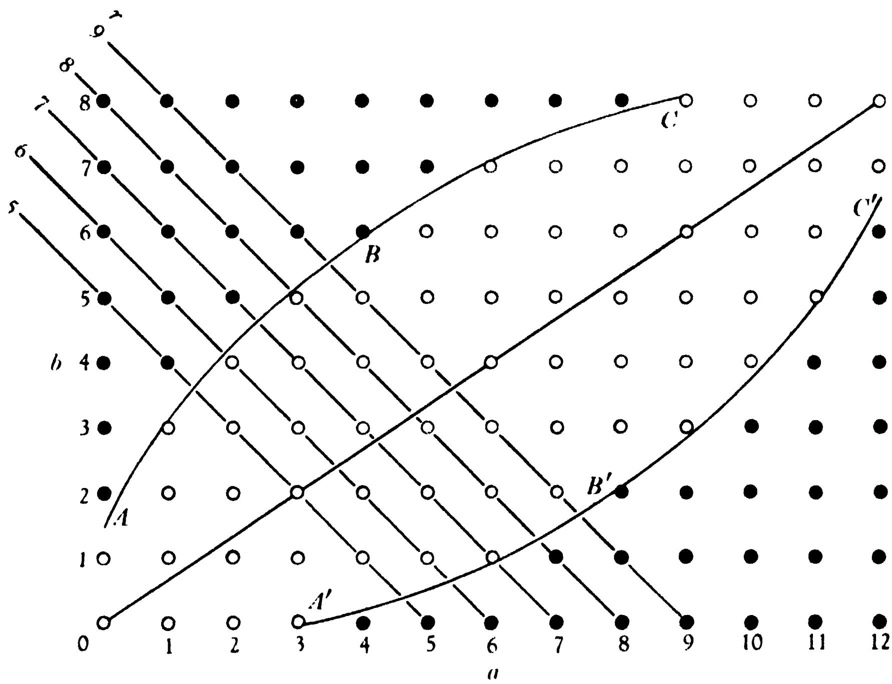


Fig. 3. The curves ABC and $A'B'C'$ represent the significance contours L_ε and L'_ε, respectively.

In this problem there have been two applications of a random selection process, not one as for Problem I, and the experimental probability set consists of the $(m+1)(n+1)$ alternative values of the doublet (a, b) $(0 \leqslant a \leqslant m, 0 \leqslant b \leqslant n)$ which can be represented in the lattice diagram shown in Fig. 3 for the special case $m = 12$, $n = 8$. It might, of course, be argued that in the hypothetical repetition of the selection process m and n need not remain constant, but this, I think, would introduce an unnecessary complication into the probability set-up.

28. The question before us is whether the result (a, b) is consistent with the hypothesis H_0 defined in equation (7) above, or whether it suggests that either $p_1 > p_2$ or that $p_1 < p_2$. A little reflexion shows that we have no reason to reject H_0 if the point (a, b) lies near the diagonal line on which $a/m = b/n$, but, broadly speaking, are more and more likely to do so the farther the point falls from this line in the direction of the corners $(0, n)$ and $(m, 0)$ of the lattice diagram. This statement requires amplification. In defining the significance contours we may consider the following question: If H_0 is not true, what departures from equality in p_1 and p_2 do we regard it of equal importance to detect? Should the power of the test be roughly the same for constant values, for example, of

$$(a)\quad p_1 - p_2, \qquad (b)\quad p_1/p_2 \qquad \text{or} \qquad (c)\quad \frac{p_1}{1-p_1}\Big/\frac{p_2}{1-p_2}?$$

The procedure which I have adopted in the sections which follow is frankly one of expediency. I have not considered in detail how to choose a family of significance contours satisfying requirements formulated in advance, but have taken those suggested by the customary large-sample procedure which gives contours of the form ABC, $A'B'C'$ drawn in Fig. 3. These will, I believe, make the power of the test to detect a difference more nearly dependent on the ratio of the odds given by (c) than on either of the expressions (a) or (b). E. B. Wilson (1941) chooses the expression (a). This point, however, needs further investigation. It should be noted that a similar problem, in the case where the sampling distributions follow the Poisson law, was discussed very fully by Przyborowski & Wilenski (1939).

29. Besides involving a 2-dimensional instead of a 1-dimensional experimental probability set, Problem II differs from Problem I in that we need an answer which is independent of the unknown common probability p of the null hypothesis. In Problem I the part of p was played by the fraction r/N given by the data. We are concerned now with what Neyman and I (Neyman & Pearson, 1933) have termed a composite hypothesis, and were it possible would like the contour levels to bound regions which are 'similar to the sample space with regard to the parameter p' (loc. cit. p. 313) (i.e. are independent of p). The following considerations show the lines along which a first attack of the problem can proceed.

30. If H_0 is true and equation (7) holds, then the probability of the observed result may be written*

$$P_2\{a \mid p, m\} \times P_2\{b \mid p, n\} = \frac{m!}{a!\,c!} p^a (1-p)^c \times \frac{n!}{b!\,d!} p^b (1-p)^d \tag{8·1}$$

$$= \frac{N!}{r!\,s!} p^r (1-p)^s \times \frac{m!\,n!\,r!\,s!}{a!\,b!\,c!\,d!\,N!} \tag{8·2}$$

$$= P_2\{r \mid p, N\} \times P_1\{a \mid N, r, m\}. \tag{8·3}$$

Thus the probability of obtaining the doublet (a, b) in sampling from two populations with a common p may be regarded as the product of two terms:

(i) The probability that $a + b = r$ or that the point (a, b) in Fig. 3 falls on a diagonal line on which r = constant. This probability, $P_2\{r \mid p, N\}$, is the $(r+1)$th term in the expansion of the binomial

$$((1-p)+p)^N.$$

(ii) The relative probability, given r, of the observed partition into a and $b = r - a$; this is independent of p and is identical with the expression $P_1\{a \mid N, r, m\}$ of equation (1), i.e. is proportional to a term of the hypergeometric series (2).

* It will be seen that $P_1\{\ \}$ has been used to denote a hypergeometric probability and $P_2\{\ \}$ a binomial probability.

31. If, now, it were possible to draw a boundary line L_ϵ such as ABC shown in Fig. 3, cutting off at the end of each diagonal, r = constant, a group of points $(a, r-a)$ such that

$$\sum_a [P_1\{a \mid N, r, m\}] = \epsilon, \tag{9}$$

where ϵ is a fraction between 0 and 1 chosen at will, then the requirement of Step 3 would be satisfied. For in rejecting H_0 when (a, b) fall beyond this boundary,* the chance of doing so if H_0 were true would be

$$\sum_{r=0}^{N} [P_2\{r \mid p, N\} \times \epsilon] = \epsilon \times \sum_{r=0}^{N} [P_2\{r \mid p, N\}] = \epsilon, \tag{10}$$

i.e. would be independent of the unknown common p of the hypothesis tested. The test would then be analogous to 'Student's' test for the significance of the difference between two means, where we have a system of contour levels L_ϵ each associated with a chance ϵ, independent of the values of any unknown parameters which are irrelevant to the composite hypothesis tested.

32. Unfortunately, this objective cannot be achieved because we are not dealing with continuous probability distributions and $P_1\{a \mid N, r, m\}$ exists only at discrete, integral values of a. If we follow the present line of approach, all that is possible is to take contour or significance levels which cut off from an end of each diagonal, r = constant, a group of points for which

$$\sum_a [P_1\{a \mid N, r, m\}] = \beta_r \leqslant \epsilon. \tag{11}$$

Then, in rejecting H_0 when (a, b) falls beyond such a contour, we know that the chance of doing so, if H_0 is true, will be

$$\sum_{r=0}^{N} [P_2\{r \mid p, N\} \times \beta_r] \leqslant \epsilon. \tag{12}$$

It is clear that the amount by which the probability falls below ϵ will be a function of p, and that in taking Step 3 we are only associating with each significance level L_ϵ an upper limit, ϵ, to the probability of rejecting H_0 when it is true.

33. We have still, of course, to determine the most appropriate system of significance levels and to set out a ready means of finding an upper limit, ϵ, associated with the level on which an observed doublet (a, b) falls.† Mr Barnard has broken new ground in

(i) defining for this Problem II one systematic method of determining a family of levels L_ϵ based on certain clearly defined principles;

(ii) determining the true upper bound to the associated probability ϵ which, in the case of small samples at any rate, may be considerably below that which has hitherto been used.

Since, however, much tabling is needed before his theoretical advance can be followed by a practical working rule available for samples of any sizes, m and n, I think it is worth while describing the cruder handling of the lattice diagram which I had discussed in 1938–9

* There would be a similar series of boundaries, L'_ϵ, below the diagonal $a/m = b/n$, such as $A'B'C'$ of Fig. 3.

† The likelihood ratio λ might be used in determining the family of significance contours, as was suggested in connexion with the general χ^2 problem (Neyman & Pearson, 1928*b*, p. 283). In large samples λ would approximately equal $e^{-\frac{1}{2}u^2}$, where u is given by equation (22) below.

lectures. This involves, perhaps, not much more than a restatement of what may be termed the classical approach to Problem II (see paras. 43 and 44 below), but it does bring out the difference between Problems I and II, which I think important.

34. It may be well to emphasize here that this distinction between the handling of Problems I and II is not universally accepted. Fisher has set out his approach as follows in a paper read before the Royal Statistical Society (1935): 'To the many methods of treatment hitherto suggested for the 2×2 table the concept of ancillary information suggests this new one. Let us blot out the contents of the table, leaving only the marginal frequencies. If it be admitted that these marginal frequencies by themselves supply no information on the point at issue, namely, as to the proportionality of the frequencies in the body of the table, we may recognize the information they supply as wholly ancillary; and therefore recognize that we are concerned only with the relative probabilities of occurrence of the different ways in which the table can be filled in, subject to these marginal frequencies.'

This view has also been supported by Yates (1934). As I understand it, Fisher would refer the observation (a, b) to a linear set (as in my Problem I), however the data have been collected; this attitude follows readily if we discard the requirement that the probability distribution used in the test must be related to the frequency distribution that would be generated by repeated application of the random sampling process employed in the experiment. It will be seen that with Fisher's approach there is a gain in simplicity in handling the analysis; it must remain a matter of opinion whether there is a loss in the relevance of the probability construct to the question at issue. It is, of course, only when handling small samples or in cases where (a, b) lies close to one of the corners $(0, 0)$ or (m, n) of the lattice that this need for choice between probability constructs is thrust upon us.

(vi) Solution of Problem II, using the normal approximation

35. If the samples are large, the calculation of hypergeometric terms becomes laborious and we turn naturally, as in so many other statistical problems, to the approximation using the normal curve. In fact, except when r or s are very small or m and n very different in magnitude, the normal curve with mean and standard deviation given by equations (3) and (4) provides a surprisingly good approximation to the relative probability distribution of a for fixed r, viz. $P_1\{a \mid N, r, m\}$ (see Appendix). Define u_ϵ as the deviate of the standardized normal curve for which

$$\epsilon = \int_{u_\epsilon}^{\infty} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}u^2} du \quad (\epsilon \leqslant \tfrac{1}{2}). \tag{13}$$

Then we can draw across the lattice diagram a significance level L_ϵ above and another L'_ϵ below* the diagonal $a/m = b/n$ such that

(i) all points (a, b) for which

$$\frac{(a+\frac{1}{2})-\bar{a}}{\sigma_a} \leqslant -u_\epsilon \tag{14}$$

lie beyond, i.e. above, L_ϵ;

(ii) and all points (a, b) for which

$$\frac{(a-\frac{1}{2})-\bar{a}}{\sigma_a} \geqslant u_\epsilon \tag{15}$$

lie beyond, i.e. below, L'_ϵ.

* The words 'above' and 'below' are used in the sense of Figs. 3 and 4.

If we wish to take special action either when a/m is significantly less than b/n or significantly greater, then we shall use both levels L_ϵ and L'_ϵ; if only, however, when $a/m < b/n$, then we use L_ϵ. The corresponding probability levels would be obtained by making ϵ for the second case twice its value for the first. Fig. 4 shows the 247 relative probabilities $P_1\{a \mid N, r, m\}$ for the case $m = 18$, $n = 12$. The unbroken, stepped lines are two contour levels determined in this way. Purely for convenience in drawing, the level with $\epsilon = 0{\cdot}05$ and $u_{0\cdot05} = 1{\cdot}6445$ has been put above the diagonal and that with $\epsilon = 0{\cdot}01$ and $u_{0\cdot01} = 2{\cdot}3263$ below.

36. If the normal approximation to the hypergeometric series were correct, it would follow that along every diagonal, r = constant, the sum of the relative probabilities for points above L_ϵ would satisfy the inequality (11). Hence the inequality (12) for the complete area of the lattice above L_ϵ would hold, whatever the value of the common p. A similar result would hold for the area below L'_ϵ. Of course, the normal approximation will not hold precisely, particularly when r or s are small, but here we shall generally be on the safe side, in the sense that the hypergeometric distribution is flat-topped with abrupt ends so that the β_r of equation (11) will be considerably less than ϵ, and often zero.

37. It is interesting to examine the results set out in Fig. 4 with the help of the detailed calculations given in Table 4. Columns (2) and (3) give, for constant r, the mean and standard deviation of $P_1\{a \mid 30, r, 18\}$, while columns (4) (for $L_{0\cdot05}$) and (8) (for $L'_{0\cdot01}$) give the cut-off points defined by the normal approximation, i.e.

$$a_1 = \bar{a} - \tfrac{1}{2} - u_{0\cdot05} \times \sigma_a \quad \text{and} \quad a_2 = \bar{a} + \tfrac{1}{2} + u_{0\cdot01} \times \sigma_a. \tag{16}$$

The sums of the relative probabilities $P_1\{a \mid 30, r, 18\}$ for $a \leqslant a_1$ and $a \geqslant a_2$ are given in cols. (5) and (9) respectively. Thus, for example, for $r = 7$

$$a_1 = 4{\cdot}2 - 0{\cdot}5 - 1{\cdot}6449 \times 1{\cdot}1543 = 1{\cdot}80,$$

and the sum of the probabilities for $a = 0$ and 1 is

$$0{\cdot}0004 + 0{\cdot}0082 = 0{\cdot}0086.$$

These are the tail sums, termed β_r in equation (11). It is clear from an examination of cols. (5) and (9) that they are all less, and many of them very much less than 0·05 and 0·01. This is inevitable with a discrete distribution containing few terms. The contour levels have been drawn conventionally in Fig. 4 as steps passing through the half-integer points and not through the cut-off points of cols. (4) and (8). Clearly, whichever way they are drawn, they will separate off the same subset of the $(m+1)(n+1)$ points in the lattice diagram.

38. The next question is this. If we were to use either of these levels, what in fact would be the chance of the sample doublet (a, b) falling beyond, if the null hypothesis were true? This will depend on the common value of p. The product sums

$$\sum_{r=0}^{N} [P_2\{r \mid p, N\} \times \beta_r] = \sum_{r=0}^{N} \left[\frac{N!}{r!\,s!} p^r (1-p)^s \times \beta_r\right] \tag{17}$$

obtained by multiplying the expressions in cols. (5) and (9) of Table 4 by the appropriate binomial terms are shown for a variety of values of p in Table 5, cols. (2) and (3). It is clear at once how far on the safe side we are in saying that these chances are $\leqslant 0{\cdot}05$ and $0{\cdot}01$ respectively. Similar calculations were carried out for a second example, taking $m = n = 10$,

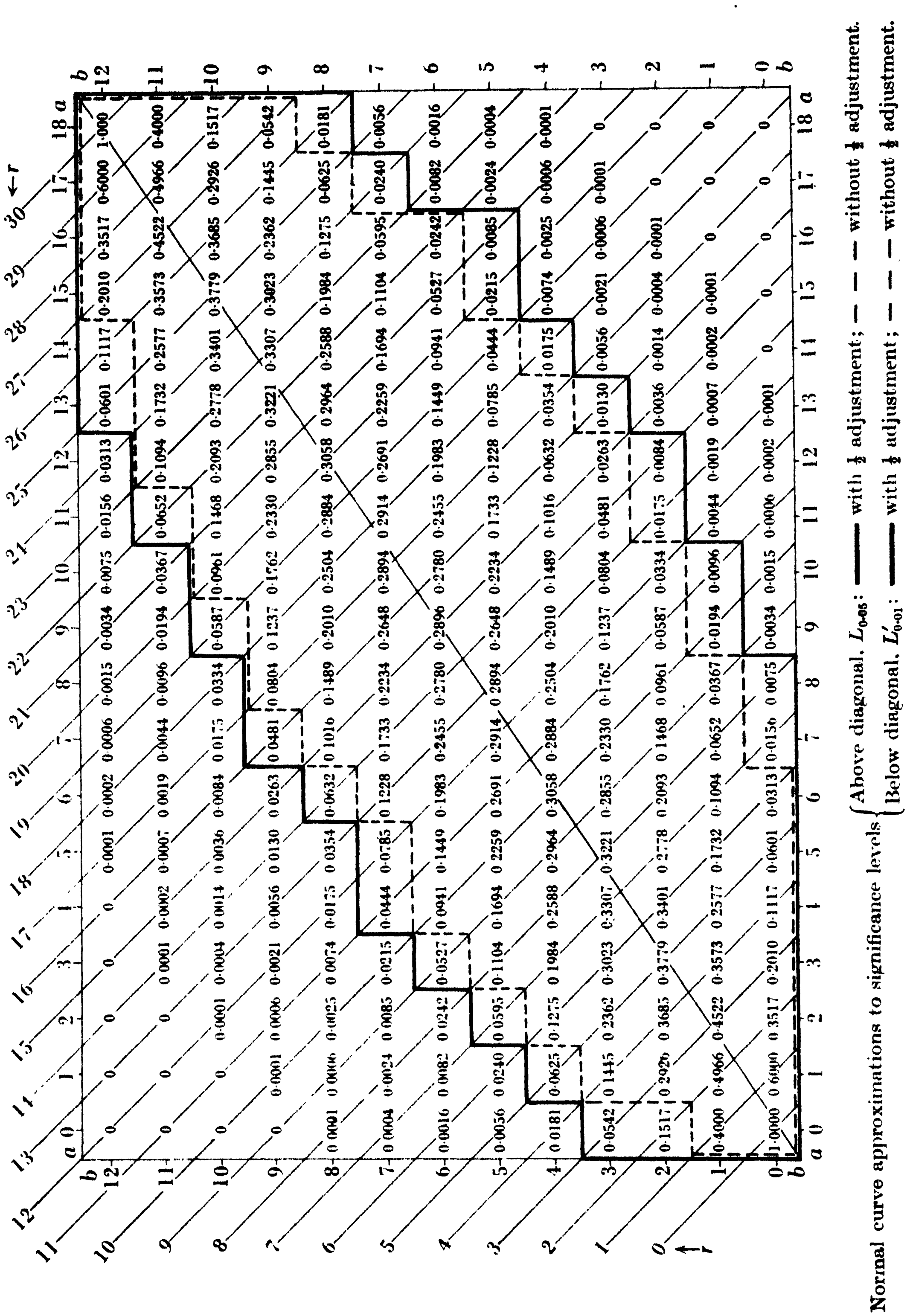


Fig. 4. Hypergeometric probabilities in lattice diagram for $m = 18$, $n = 12$.

Table 4. *Significance levels for case $m = 18$, $n = 12$*

r	$\bar{a}$	σ_a	Details for L_ϵ: $\epsilon = 0{\cdot}05$, $u_{0\cdot05} = 1{\cdot}6449$				Details for L'_ϵ: $\epsilon = 0{\cdot}01$, $u_{0\cdot01} = 2{\cdot}3263$				r
			Method 1		Method 2		Method 1		Method 2		
			Cut-off $\bar{a} - \frac{1}{2} - u_\epsilon\sigma_a$	Sum of terms beyond cut-off	Cut-off $\bar{a} - u_\epsilon\sigma_a$	Sum of terms beyond cut-off	Cut-off $\bar{a} + \frac{1}{2} + u_\epsilon\sigma_a$	Sum of terms beyond cut-off	Cut-off $\bar{a} + u_\epsilon\sigma_a$	Sum of terms beyond cut-off	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
0	0	0	−0·50	0	0	0	0·50	0	0	0	0
1	0·6	0·4899	−0·71	0	−0·21′	0	2·24	0	1·74	0	1
2	1·2	0·6808	−0·42	0	0·08	0·1517	3·28	0	2·78	0	2
3	1·8	0·8187	−0·05	0	0·45	0·0542	4·20	0	3·70	0	3
4	2·4	0·9277	0·37	0·0181	0·87	0·0181	5·06	0	4·56	0	4
5	3·0	1·0171	0·83	0·0056	1·33	0·0681	5·87	0	5·37	0	5
6	3·6	1·0917	1·30	0·0256	1·80	0·0256	6·64	0	6·14	0	6
7	4·2	1·1543	1·80	0·0086	2·30	0·0681	7·39	0	6·89	0·0156	7
8	4·8	1·2069	2·31	0·0267	2·81	0·0267	8·11	0	7·61	0·0075	8
9	5·4	1·2507	2·84	0·0091	3·34	0·0618	8·81	0·0034	8·31	0·0034	9
10	6·0	1·2865	3·38	0·0241	3·88	0·0241	9·49	0·0015	8·99	0·0209	10
11	6·6	1·3152	3·94	0·0080	4·44	0·0524	10·16	0·0006	9·66	0·0102	11
12	7·2	1·3370	4·50	0·0197	5·00+	0·0982	10·81	0·0046	10·31	0·0046	12
13	7·8	1·3524	5·08	0·0414	5·58	0·0414	11·45	0·0020	10·95	0·0195	13
14	8·4	1·3615	5·66	0·0145	6·16	0·0777	12·07	0·0007	11·57	0·0091	14
15	9·0	1·3646	6·26	0·0301	6·76	0·0301	12·67	0·0038	12·17	0·0038	15
16	9·6	1·3615	6·86	0·0091	7·36	0·0572	13·27	0·0015	12·77	0·0145	16
17	10·2	1·3524	7·48	0·0195	7·98	0·0195	13·85	0·0060	13·35	0·0060	17
18	10·8	1·3370	8·10	0·0380	8·60	0·0380	14·41	0·0022	13·91	0·0197	18
19	11·4	1·3152	8·74	0·0102	9·24	0·0689	14·96	0·0080	14·46	0·0080	19
20	12·0	1·2865	9·38	0·0209	9·88	0·0209	15·49	0·0026	14·99	0·0241	20
21	12·6	1·2507	10·04	0·0401	10·54	0·0401	16·01	0·0006	15·51	0·0091	21
22	13·2	1·2069	10·71	0·0075	11·21	0·0727	16·51	0·0025	16·01	0·0025	22
23	13·8	1·1543	11·40	0·0156	11·90	0·0156	16·99	0·0086	16·49	0·0086	23
24	14·4	1·0917	12·10	0·0313	12·60	0·0313	17·44	0·0016	16·94	0·0256	24
25	15·0	1·0171	12·63	0	13·13	0·0601	17·87	0·0056	17·37	0·0056	25
26	15·6	0·9277	13·57	0	14·07	0·1117	18·26	0	17·76	0·0181	26
27	16·2	0·8187	14·35	0	14·85	0	18·60	0	18·10	0	27
28	16·8	0·6808	15·18	0	15·68	0	18·88	0	18·38	0	28
29	17·4	0·4899	16·09	0	16·59	0	19·04	0	18·54	0	29
30	18·0	0	17·50	0	18·00	0	18·50	0	18·00	0	30

and the results are shown in Table 5, cols. (6) and (7). In this case, the actual chances of (a, b) falling on or beyond the significance levels are even further below the nominal limits of 0·05 and 0·01. In fact, it becomes clear that in the case of small samples, at any rate, this method of introducing the normal approximation gives such an overestimate of the true chances of falling beyond a contour as to be almost valueless.

Table 5. *Showing the difference between nominal and actual significance levels*

p (if H_0 true)	1st example: $m = 18$, $n = 12$				2nd example: $m = 10 = n$				p (if H_0 true)
	Method 1		Method 2		Method 1		Method 2		
	True chance of falling on or beyond		True chance of falling on or beyond		True chance of falling on or beyond		True chance of falling on or beyond		
	$L_{0\cdot05}$	$L'_{0\cdot01}$	$L_{0\cdot05}$	$L'_{0\cdot01}$	$L_{0\cdot05}$	$L'_{0\cdot01}$	$L_{0\cdot05}$	$L'_{0\cdot01}$	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0·05	0·0010	0·0000	0·0478	0·0000	0·0000	0·0000	0·0069	0·0000	0·05
0·1	0·0054	0·0000	0·0602	0·0003	0·0005	0·0000	0·0251	0·0005	0·1
0·2	0·0141	0·0003	0·0483	0·0043	0·0037	0·0007	0·0455	0·0037	0·2
0·3	0·0174	0·0012	0·0490	0·0091	0·0058	0·0014	0·0495	0·0058	0·3
0·4	0·0204	0·0023	0·0542	0·0108	0·0062	0·0017	0·0546	0·0062	0·4
0·5	0·0219	0·0028	0·0498	0·0109	0·0062	0·0015	0·0572	0·0062	0·5
0·6	0·0221	0·0035	0·0437	0·0119	Repeat as for $1-p$				0·6
0·7	0·0204	0·0037	0·0431	0·0120					0·7
0·8	0·0126	0·0031	0·0459	0·0113					0·8
0·9	0·0019	0·0009	0·0282	0·0052					0·9
0·95	0·0001	0·0001	0·0058	0·0010					0·95

39. Before considering a second method, it will be useful to recapitulate certain characteristics of what I have termed Method 1. It provides for any nominal value of ϵ one systematic procedure of defining a critical boundary or significance level cutting off a region from the lattice diagram. Neither the subgroup of points cut off, nor the sum of the probabilities associated with them for a given p, will alter continuously with ϵ; they will change by discrete steps as the cut-off point, defined in para. 37, passes through a point (a, b). While we shall sometimes want to know whether the observed (a, b) falls beyond a level L_ϵ specified in advance, more often we shall ask what is the level on which (a, b) falls. This, using Method 1, we find by calculating

$$u = \frac{\bar{a}-(a+\frac{1}{2})}{\sigma_a} \quad \text{if} \quad a<\bar{a} \qquad \text{or} \qquad u = \frac{a-\frac{1}{2}-\bar{a}}{\sigma_a} \quad \text{if} \quad a>\bar{a}, \tag{18}$$

and finding ϵ from the normal integral of equation (13). In this way the nominal chance ϵ will be a little nearer the true upper limit than the figures in Table 5 suggest,* but not enough to modify the criticism expressed above.

* It will be seen from Table 4 that no point (a, b) gives a β_r in cols. (5) and (9) of exactly 0·05 or 0·01, respectively, so that no points actually lie on $L_{0\cdot05}$ or $L_{0\cdot01}$.

40. *Method* 2. The introduction of the correction of $\frac{1}{2}$ for continuity is certainly appropriate in using the normal approximation to the hypergeometric series in Problem I, but I think it is not helpful in Problem II where we are concerned with a 2-dimensional experimental probability set. If instead of obtaining significance levels L_ϵ and L'_ϵ as in paras. 35–37, we obtain them from inequalities similar to (14) and (15) but with the correction of $\frac{1}{2}$ omitted, then there are several points to be noted:

(*a*) For the significance level L_ϵ, the expression

$$\beta_r = \sum_a [P_1\{a \mid N, r, m\}], \tag{19}$$

where the summation is for values of a on the diagonal, r = constant, for which

$$a \leqslant a_1 = \alpha - u_\epsilon \times \sigma_a \tag{20}$$

will be sometimes less and sometimes greater than ϵ. Hence, in the balance, it seems likely that the chance of the point (a, b) lying beyond L_ϵ or

$$\sum_{r=0}^{N} \left[\frac{N!}{r!\,s!} p^r (1-p)^s \times \beta_r\right] \tag{21}$$

will lie closer to ϵ than when the $\frac{1}{2}$ correction is used. The position will be the same for L'_ϵ.

(*b*) In drawing repeated samples of m and n from two populations in which there is a common chance, p, of an individual possessing character A, the ratio

$$u = \frac{a-\alpha}{\sigma_a} = \frac{a - rm/N}{\sqrt{\dfrac{mnrs}{N^2(N-1)}}} \tag{22}$$

has, whatever be p, (i) an expectation of zero, (ii) a unit standard deviation.* The shape of the distribution will, of course, depend on p, but, *faut de mieux*, we may not in the long run do too badly by assuming it to be normal. It is, of course, the weighted combination of a number of hypergeometric series whose shape depends on r.

41. Consider the result of applying this Method 2 to the case $m = 18$, $n = 12$ already discussed. The procedure for determining the 0·05 and 0·01 significance levels will be exactly as under Method 1, except that the continuity correction of $\frac{1}{2}$ is omitted. The resulting levels are shown as dashed, stepped lines in Fig. 4.† They fall, on the whole, inside the significance levels obtained by Method 1. Now turn to Table 4, where cols. (6) and (10) show the cut-off points a half unit further in towards the diagonal $a/m = b/n$. Cols. (7) and (11) give the values of β_r; some of these are considerably above the nominal values of $\epsilon = 0{\cdot}05$ and $0{\cdot}01$, others are still well below. But from the approach to Problem II that has been adopted, this is immaterial since the experimental probability set is the 2-dimensioned one of the lattice diagram and is not restricted to the diagonal r = constant on which the observed point (a, b) may happen to lie. What we are concerned with is the summed chance given by expression (21) and the value of this is given for eleven values of p in cols. (4) and (5) of Table 5. It will be seen that this true chance does sometimes exceed the nominal values of 0·05 and 0·01,

* Provided cases where r or s are zero, making the expression (22) indeterminate with $u = 0/0$, are excluded. Mr Barnard has pointed out that one way of avoiding this exclusion would be to lay down that, when $u = 0/0$, we assign to the ratio a value chosen at random from a population (say normal) with zero mean and unit variance.

† Again, for convenience the 5 % level is drawn above and the 1 % level below the diagonal.

but never by very much. Again, for the second example with $m = 10 = n$ (Table 5, cols. (8) and (9)) the true chance, while it sometimes exceeds the nominal value, is always considerably nearer it than using the significance levels of Method 1.

42. It is clear that no final conclusions can be based on two numerical examples, but it seems that the test of the null hypothesis in Problem II should be carried out as follows:

(*a*) When m, n, r or s are small, with the help of tables prepared on Barnard's lines, based on an ordered classification of the points in the lattice diagram, and giving the true upper bound of the chance that a point (a, b) falls on or beyond the level on which the observed result lies. The particular basis of his classification may, of course, be modified.

(*b*) When m, n, r and s are large, by assuming that the u of equation (22) is a normal deviate with unit standard deviation.

(vii) The classical approach to Problem II

43. It has recently become customary to regard the test of significance applied to data given in a 2×2 table as the limiting case of a χ^2 test with one degree of freedom. But Problem II was originally answered in somewhat different terms. It was noted that if

$$p_1(A) = p_2(A) = p, \tag{23}$$

then the fractions a/m and b/n would both have expectations of p and variances of $p(1-p)/m$ and $p(1-p)/n$, respectively. Hence, if the null hypothesis were true, the difference

$$d = \frac{a}{m} - \frac{b}{n} \tag{24}$$

would have

$$\left.\begin{aligned} \text{mean } d &= 0 \\ \sigma_d &= \sqrt{\left[p(1-p)\left(\frac{1}{m}+\frac{1}{n}\right)\right]} \end{aligned}\right\}. \tag{25}$$

In large samples, therefore, it might be expected that

$$\frac{d}{\sigma_d} = \frac{a/m - b/n}{\sqrt{[p(1-p)(1/m+1/n)]}} \tag{26}$$

would be approximately normally distributed. Since by the nature of the problem the common value of p was unknown, an estimate was made from the sample, namely,

$$\hat{p} = \frac{a+b}{m+n} = \frac{r}{N}. \tag{27}$$

Substituting this into equation (26), we have

$$\frac{d}{s_d} = \frac{a/m - b/n}{\sqrt{[(r/N)(1-r/N)(1/m+1/n)]}} \tag{28·1}$$

$$= \frac{a - rm/N}{\sqrt{\left(\dfrac{mnrs}{N^3}\right)}}. \tag{28·2}$$

44. The form (28·2) is easily derived from (28·1), if we remember that $b = r - a$, $s = N - r$ and $m + n = N$.* It is seen that the ratio d/s_d is identical with the ratio u of equation (22), except for a factor $\sqrt{[(N-1)/N]}$ which is unimportant in large samples. Thus the classical test is practically identical with that suggested in paras. 40–42 above, though the two tests are differently derived.

* A third alternative form is, of course, $(ad - bc)\sqrt{N}/\sqrt{(mnrs)}$.

(viii) Problem III

45. This may be described as the test for the independence of two characters A and B. It is supposed that the probability that an individual selected at random will possess character A is $p(A)$ and that he will not possess it is $p(\bar{A}) = 1-p(A)$. The corresponding probabilities for character B are $p(B)$ and $p(\bar{B}) = 1-p(B)$. Four alternative combinations of the characters may occur, which may be denoted by AB, $A\bar{B}$, $\bar{A}B$ and $\bar{A}\bar{B}$. The various probabilities are set out in Table 6A. If the null hypothesis, H_0, specifying the independence of A and B is true, then

$$p(AB) = p(A)\times p(B), \quad p(A\bar{B}) = p(A)p(\bar{B}), \quad \text{etc.} \tag{29}$$

To test the hypothesis, we have a random sample of N observations with frequencies of occurrence of the combinations AB, $A\bar{B}$, etc., which may be classified in the 2×2 scheme of Table 6B. The sampling conditions are such that the probabilities of Table 6A are the same for all individuals selected, or, in conventional terms, the sample is drawn from an infinite population. Barnard calls this problem that of the double dichotomy.

Table 6A. *Probabilities*

	A	$\bar{A}$	Total
B	$p(AB)$	$p(\bar{A}B)$	$p(B)$
$\bar{B}$	$p(A\bar{B})$	$p(\bar{A}\bar{B})$	$p(\bar{B})$
Total	$p(A)$	$p(\bar{A})$	1

Table 6B. *Sample data*

	A	$\bar{A}$	Total
B	a	c	m
$\bar{B}$	b	d	n
Total	r	s	N

46. In Problem III there is only one application of a random process, the selection of N individuals, each one of which must fall into one or other of four alternative categories. If the random process were repeated and another sample of N drawn, not only are the frequencies a, b, c and d free to vary, but also *both* marginal totals, i.e. m may change as well as r. The experimental probability set will therefore contain results (a, b, c, d) restricted by the conditions (i) that none of the frequencies can be negative and (ii) that

$$a+b+c+d = N. \tag{30}$$

Geometrically, as Barnard points out, the set can be represented in 3 dimensions by points at unit intervals within a tetrahedron obtained by placing on top of one another the series of 2-dimensioned lattices of dimensions

$$0\times n, \quad 1\times(n-1), \quad 2\times(n-2), \quad \ldots, \quad (m-1)\times 1, \quad m\times 0. \tag{31}$$

47. We are again testing a composite hypothesis and should like to determine a family of critical surfaces to be used as significance levels, dividing the points within the tetrahedron in such a way that the chance of the sample point (a, b, c, d)* lying outside a given surface L_ϵ is equal to ϵ, whatever the values of the unknown probabilities $p(A)$ and $p(B)$. But again, as in Problem II, owing to the discontinuity in the set of points, there are no 'similar

* In view of the condition (30), the point can be defined by three co-ordinates, e.g. as (a, b, c), (a, b, m) or (a, r, m). In view of the form of equation (32), the last system of co-ordinates will be used.

regions'. We note that if H_0 is true, the probability of the observed result is a term of the multinomial expansion, viz.

$$\frac{N!}{a!\,b!\,c!\,d!}p(AB)^a p(A\bar{B})^b p(\bar{A}B)^c p(\bar{A}\bar{B})^d$$

$$= \frac{N!}{a!\,b!\,c!\,d!}p(A)^{a+b} p(B)^{a+c} p(\bar{A})^{c+d} p(\bar{B})^{b+d}$$

$$= \frac{N!}{m!\,n!}p(B)^m (1-p(B))^n \times \frac{N!}{r!\,s!}p(A)^r (1-p(A))^s \times \frac{m!\,n!\,r!\,s!}{a!\,b!\,c!\,d!\,N!}$$

$$= P_2\{m \mid p(B), N\} \times P_2\{r \mid p(A), N\} \times P_1\{a \mid N, r, m\}. \qquad (32)$$

Here, the notation of para. 30 has been repeated.

48. Thus the probability of obtaining a sample represented by the triplet (a, r, m) may be regarded, if the characters A and B are independent, as the product of three terms:

(i) The probability of drawing m individuals with character B in a random sample of N, i.e. the probability that (a, r, m) falls in a horizontal section of the tetrahedron on which m = constant. This is the $(m+1)$th term in the expansion of the binomial

$$\{(1-p(B))+p(B)\}^N.$$

(ii) The probability of drawing r individuals with character A in a random sample of N, i.e. the probability that (a, r, m) falls on the vertical section of the tetrahedron on which r = constant. This is the $(r+1)$th term in the expansion of

$$\{(1-p(A))+p(A)\}^N.$$

(iii) The probability, given m and r, of the observed partition within the 2×2 table. This term represents the relative probability associated with the points lying along a straight line m = constant, r = constant; it is, of course, the same expression as has arisen in Problems I and II and is proportional to a term in the hypergeometric series $F(-r, -m, n-r+1, 1)$.

49. We are faced with a situation similar to that met under Problem II. Were it possible to cut off from each line on which m = constant, r = constant, a group of points such that

$$\sum_a [P_1\{a \mid N, r, m\}] = \epsilon, \qquad (33)$$

then the subset of points within the tetrahedron composed of the sum of these groups for all possible combinations of m and r would have the property required of a 'critical region' in a significance test: i.e. the chance that the point (a, r, m) is included in the region, if H_0 is true, would be ϵ whatever values the irrelevant probabilities $p(A)$ and $p(B)$ assumed. However, (33) cannot be satisfied in general, and all that is possible is to define a family of significance contours such that the chance of a sample point falling beyond any one of them, say L_ϵ, is $\leqslant \epsilon$. By using the normal approximation to the sum of the hypergeometric tail-terms with the correction for continuity as described in paras. 35–39 for Problem II, we shall be very much on the safe side, i.e. the formal level of ϵ is likely to be much above the true chance of falling beyond the level, whatever be $p(A)$ or $p(B)$. The presence of the two binomial terms in equation (32) instead of the single term in equation (8·3), makes it likely that the overestimation of ϵ will be greater in Problem III than in II. It is to be expected, therefore, that any any rate when neither m, n, r or s are too small, the better approximation will be obtained by referring the u of equation (22) to the normal probability scale.

50. The handling of Problem III is discussed briefly by Barnard on p. 136 above. There is clearly room for further investigation. The general nature of the approximation

involved is of course that which arises in every χ^2 test for goodness of fit or for independence in an $h \times k$ table, where we replace a distribution consisting of a finite set of probabilities at discrete points in multiple space by a continuous distribution for which integration outside ellipsoidal contours is straightforward.

(ix) General comment

51. The duties of the statistician lie at many levels. He may be required merely to apply an established technique of analysis to an assembly of numerical data and this application may result in a statement, based on probability theory, of a 'level of significance' or a 'confidence interval', which will be used by others. Or he may be called on to share in planning the investigation or experiment which is to provide the data and then to draw conclusions from their analysis which will lead to further action. In this final role he needs to bring into play faculties which are no monopoly of his calling, the qualities of sound judgement which are the characteristics of a well trained, scientific mind. In the weighing of evidence, the result of the statistical analysis, expressed in one or more conventional probability figures, is only one factor in the summing up; as important, may be, is the question of whether the mathematical model is a fair counterpart to the happenings in the observational field. In addition, there will often be much information coming from outside the range of the immediate investigation, yet hardly expressible in numerical terms, which must influence decision.

52. It is perhaps hard experience gained in certain fields of war-time research, where decisions had to be reached on statistical data far less ample than could be wished, which has forced my own attention to this question: What weight do we actually give to the precise value of a probability measure when reaching decisions of first importance? One subject for examination falling under this inquiry is clearly the logical basis of the reasoning process by which judgement is influenced as a result of the application of a test of significance. This was the theme on which this paper opened. The approach illustrated in the pages which followed is a personal one and is set down, with no claim to be the best, in order to provoke thought and discussion. There appears no short route to a right answer in this matter; each individual who hopes to use his own judgement to the full in drawing conclusions from the statistical analysis of sampling data, must decide for himself what he requires of probability theory.

53. In the approach which I have followed and illustrated on the analysis of data classed in a 2×2 table, the appropriate probability set-up is defined by the nature of the random process actually used in the collection of the data. Consideration of this point forms the initial step in the determination of the appropriate test. On this score, what I have termed Problems I, II and III are differentiated. The difference is fundamental and lies at the bottom of the dilemma to which the Barnard-Fisher correspondence in *Nature* drew attention. It can be illustrated on the following data, given in Table 7, where I shall suppose that the effect we are interested in is that making a significantly greater than b.

54. If (a) the results have been obtained by random assignment of Treatment 1 to eighteen out of thirty individuals and Treatment 2 to the remaining twelve, and

(b) we merely ask whether the results are consistent with the hypothesis that the treatments are equivalent as far as these thirty individuals are concerned, so that the difference between the proportions 15/18 and 5/12 may reasonably be ascribed to a chance fluctuation,

(*c*) we are then concerned with Problem I, i.e. simply with the probabilities associated with the points $(a, 20-a)$ on the diagonal $r = 20$ of Fig. 4. The chance of getting $a \geqslant 15$, if the null hypothesis is true, is 0·0241,* or, using a common phrase, we can speak of the result being significant at the 2·5 % level.

55. On the other hand, if a sample of 18 has been drawn randomly from one population and a sample of 12 independently from a second and we wish to test whether $p_1(A) = p_2(A)$, then it seems to be an artificial procedure to restrict the experimental probability set to the 11 points on the line $r = 20$, i.e. to the values of a: 8, 9, ..., 18. A repetition of the double sampling process could give us a result (a, b) falling at any of the $19 \times 13 = 247$ points in the lattice diagram of Fig. 4. There will be a number of ways of defining a family of significance levels for this 2-dimensioned set; if we adopt that discussed in paras. 40–41, which

Table 7

For problem I	For problem II	Frequency of results		Total
		A	$\bar{A}$	
1st treatment	Sample from 1st population	$a = 15$	$c = 3$	$m = 18$
2nd treatment	Sample from 2nd population	$b = 5$	$d = 7$	$n = 12$
	Total	$r = 20$	$s = 10$	$N = 30$

gives as two of its members the dotted, stepped lines shown in Fig. 4, we can say that the chance of a result falling beyond the lower line is certainly less than 0·015.† The observed point, with $a = 15$, $b = 5$ falls beyond the line, so that the result is undoubtedly 'significant at the 1·5 % level'.

56. These two probabilities, 2·5 and 1·5 %, are not the same, but there is no inconsistency in their difference. The character of the two investigations is different and to treat Problem II as though it were Problem I seems to call for a probability set-up which is unnecessarily artificial, when a simpler one is available. Admittedly by getting what seems to me a closer relation between the probability set-up and the experimental procedure, we have sacrificed some simplicity in handling the 2×2 table. But this is only the case when dealing with small numbers. For large numbers the methods of handling Problems I, II and III become, practically, identical.

57. Consider again the heavy shell problem described in para. 7 above. If we are to introduce probability theory, it seems to me that we should regard the problem as one in which we have a sample of $m = 12$ from the possible output of shell made to one design or by one firm and of $n = 8$ from the possible output of a second. This sampling may be hypothetical in that these may be 'pilot' shell, the first off production; nevertheless, this construct is

* For the normal curve approximation, using the correction for continuity, we find

$$u = (15 - \tfrac{1}{2} - 12{\cdot}0)/1{\cdot}2865 = 1{\cdot}943.$$

The proportionate area under the normal curve beyond this deviation is 0·026.

† Table 5, col. (5) shows the largest value of this chance to be 0·0120 for $p = 0{\cdot}3$. This figure cannot be much exceeded for other p's though I have not determined the precise maximum. I give 0·015 as a safe-side limit.

clearly less artificial than one in which, on the null hypothesis, we regard the experiment as though it were made on twenty shells, to twelve of which has been randomly assigned the label 'Made by firm X' and to the other eight, 'Made by firm Y'.

58. It is clear that in the heavy shell problem there may be many reasons to doubt whether the rounds fired can be regarded as a random sample from future output. That is why I have emphasized that the exploration which the statistician makes in private will not necessarily be presented in figures at the conference table. In this example, the proportions of successful perforations were 2/12 and 5/8; these put us on the line, $r = 7$, of the lattice diagram for which the hypergeometric probabilities were shown in Fig. 2. The sum of the terms with $a \leqslant 2$ is 5·2 % (normal approximation, using the $\frac{1}{2}$-correction, 5·6 %). This is the chance of getting as great or a greater positive difference, $b-a$, if H_0 were true, treating the case as Problem I. Barnard's method has not yet been extended to cover this case, but if we were to use the large sample method for handling Problem II, described in my paras. 40–41, we should find from equation (22) that

$$u = (2-4\cdot2)/1\cdot072 = -2\cdot05,$$

which puts (a, b) outside the upper 2·5 % level.

59. Were the action taken to be decided automatically by the side of the 5 % level on which the observation point fell, it is clear that the method of analysis used would here be of vital importance. But no responsible statistician, faced with an investigation of this character, would follow an automatic probability rule. The result of either approach would raise considerable doubts as to whether the performance of the first type of shell was as good as that of the second, but without the whole background of the investigation it is impossible to say what the statistician's recommendation as to further action would be.

60. In the example of the proof of anti-tank shot discussed in para. 6, the chance of perforation, p, while varying from plate to plate and batch to batch, will almost certainly not range through the whole interval 0–1. The striking-velocity of the shot would also probably be adjusted so that for average proof-plate and batches, p was near $\frac{1}{2}$. Then the discriminating level (or levels*) set across the 13×13 lattice diagram would be fixed paying regard to the likely variation in p; thus a fairly close upper limit could be calculated to the true probability of (a, b) falling beyond the level if the fresh batch were of the same quality as the standard. This is the upper limit of the risk of segregating the batch wrongly.

61. Precisely similar problems arise for consideration in even more difficult form in the analysis of data arranged in a $h \times k$ table, where h or k or both are > 2. It has become common practice to speak of the solution of this problem in terms of 'fixed marginal totals', but it may be questioned whether the restriction in the experimental probability set implied is generally appropriate. The frequencies in a $h \times k$ table may have been obtained by many different sampling procedures for, as in the 2×2 problem, a single form of tabular presentation will follow from a variety of types of investigation. For most of these, a repetition of the random process of selection would give results with either one or both sets of marginal totals changed.

62. For convenience in solution we may, of course, start by considering the distribution of our test criterion, on the null hypothesis, within the sub-set of results for which the margins

* It is possible that two levels might be taken with the associated proof rules: (i) if (a, b) falls beyond the outer one, reject the batch; (ii) if between outer and inner, fire further rounds; (iii) if within the inner level, accept the batch.

are fixed. If this distribution were the same whatever these fixed values, then the overall, distribution for unrestricted sampling would be the same as that for variation subject to fixed margins, Thus, mathematically, the solution of the partial problem would be a step in the solution of the complete one. But when applying χ^2 analysis to an $h \times k$ table, this result is only true as a large-sample approximation.

63. If we use the mathematical model which it is suggested gives the most direct aid in reasoning from the observations, i.e. that which regards the experimental probability set as generated by a repetition of the random process of selection used in collecting the data, then in the majority of cases we cannot regard the marginal totals as fixed. Thus a rigorous treatment would lead, as in the case of the 2×2 table, to a differentiation into a number of solutions. It is to be hoped, however,* unless the numbers in the margins are very small, that the χ^2 approximation with its appropriate degrees of freedom† will give results which are not misleading. This approximation leads, of course, in the 2×2 table to the reference of the ratio u of equation (22) to the normal probability scale. Some aspects of the approximation in this more general case were discussed by Yates (1934, pp. 233–35).

64. In closing I should like again to acknowledge my indebtedness to Mr G. A. Barnard. Having had the good fortune to discuss these problems with him and see drafts of his work over a period of 2 or 3 years it is difficult to say how many of his ideas have been built unconsciously into my own earlier approach. But I am especially aware of the clarification which his emphasis on the distinction between Problems I, II and III brought to my survey. I am also very grateful to Mr M. G. Kendall, Dr R. C. Geary and Dr B. L. Welch for a number of helpful criticisms, and to Mrs Maxine Merrington for her extensive computing work, which has alone made possible the various numerical illustrations that I have given.

* From the point of view both of the exponents of the fixed marginal and unrestricted marginal approach.

† The statement that, for example, in applying the test of independence of two characters to an $h \times k$ table, the degrees of freedom are $(h-1) \times (k-1)$, does not of course mean that sampling is restricted by fixed marginal totals. All that is implied is that approximately the overall distribution of the χ^2 function of the observations used, is the same as that for sampling within the restricted sub-set; this is because the distribution within each sub-set is approximately independent of the particular marginal totals which define it.

REFERENCES

BARNARD, G. A. (1945*a*). *Nature, Lond.*, **156**, 177.
BARNARD, G. A. (1945*b*). *Nature, Lond.*, **156**, 783.
FISHER, R. A. (1935). *J. Roy. Statist. Soc.* **98**, 39.
FISHER, R. A. (1941). *Science*, **94**, 210.
FISHER, R. A. (1945*a*). *Nature, Lond.*, **156**, 388.
FISHER, R. A. (1945*b*). *Sankhyā*, **7**, 130.
KENDALL, M. G. (1943). *The Advanced Theory of Statistics*, **1**. London: Charles Griffin and Co. Ltd.
NEYMAN, J. & PEARSON, E. S. (1928*a*). *Biometrika*, **20** A, 195.
NEYMAN, J. & PEARSON, E. S. (1928*b*). *Biometrika*, **20** A, 263.
NEYMAN, J. & PEARSON, E. S. (1933). *Philos. Trans.* A, **231**, 289.
NEYMAN, J. & PEARSON, E. S. (1936). *Statist. Res. Mem.* **1**, 113.
NEYMAN, J. & PEARSON, E. S. (1938). *Statist. Res. Mem.* **2**, 25.
PEARSON, K. (1899). *Phil. Mag.* **47**, 236.
PRZYBOROWSKI, J. & WILENSKI, H. (1939). *Biometrika*, **13**, 313.
WILSON, E. B. (1941). *Science*, **93**, 557.
WILSON, E. B. (1942). *Proc. Nat. Acad. Sci., Wash.*, **28**, 94.
YATES, F. (1934). *J. Roy. Statist. Soc.* Suppl. **1**, 217.

APPENDIX

THE NORMAL CURVE APPROXIMATION IN PROBLEM I

1. The following Tables 8 and 9 (A), (B) and (C) show the order of accuracy which results from using the normal curve integral as an approximation to the tail sums in the series

$$P_1\{a \mid N, r, m\} = \frac{m!\,n!\,r!\,s!}{a!\,b!\,c!\,d!\,N!} \tag{34}$$

the terms of which are proportional to those in the hypergeometric series

$$F(-r,\ -m,\ N-m-r+1,\ 1).$$

Here a is a variable which can assume the range of positive, integral values indicated under (i), (ii) and (iii) in para. 20 above, while N, r and m are fixed. The relation between these quantities and b, c, d, n and s is given in Table 1, para. 17. The method of approximation, using the '$\frac{1}{2}$' correction for continuity, has been discussed in para. 25.

2. Table 8 takes the case of an equal partition, $m = n = \frac{1}{2}N$, and shows the sum of the terms in the expression (34) for which $a \geqslant a_1$ which is also the sum of terms for which $a \leqslant r - a_1$. For $m \neq n$, results are given in Table 9 for $m > n$ and for the following proportionate partitions of N:

(A) $m = \frac{3}{5}N$, $n = \frac{2}{5}N$; (B) $m = \frac{4}{5}N$, $n = \frac{1}{5}N$; (C) $m = \frac{9}{10}N$, $n = \frac{1}{10}N$.

Here sums of terms at both tails of the series are needed. The sums (or chances of $a \geqslant a_1$ or $\leqslant a_1$) have not been given for all possible values of a_1 but, broadly speaking, for those within the limits where significance is likely to be in question. Sums below 0·0010 have generally been omitted. In each case the true sum of the terms (34) is compared with the approximation from the normal integral.

3. In drawing conclusions from the comparison, we have to decide what degree of accuracy is called for. Clearly the normal integral does not give mathematically exact results to 4 decimal places. On the other hand, except for certain instances where the partition is very unequal ($m = \frac{4}{5}N$ and $\frac{9}{10}N$) and r is small, the order of the approximation may be said to follow that of the series closely. If decisions are made by rule of thumb, according to the side of the 5 % or 1 % significance level on which a falls, then there are a number of entries in the tables where the approximation would give a on the wrong side. But one may question whether judgement of significance based on a single experiment can in fact be made sensitive to a difference between, say, 0·06 and 0·04 (odds of 16 to 1 and 24 to 1) or between 0·012 and 0·008 (odds of 82 to 1 and 124 to 1) and, given such latitude in accuracy, the approximation will be found generally sufficient. These must be points, however, where personal opinions will differ. Whatever views are held, the tables are sufficiently extensive to make it possible to obtain from them a rough measure of the accuracy of approximation in a wide range of cases.

4. It will be noted that in the symmetrical case ($m = \frac{1}{2}N$) and also when $m = \frac{3}{5}N$ the normal approximation for the tail sum is almost invariably a little too large. Undoubtedly for the symmetrical case an improved approximation could be obtained by modifying the $\frac{1}{2}$ correction used in calculating the ratio of deviation to standard deviation. This second order term would, however, need to vary with the probability level, thus complicating the procedure,

Table 8. *Case of equal partition*, $m = n = \frac{1}{2}N$. *Chance that* $a \geqslant a_1 =$ *chance that* $a \leqslant r - a_1$

Partition		$m = n = 50$		$m = n = 30$		$m = n = 20$		$m = n = 15$		$m = n = 10$			
r	a_1	True	Normal approx.	True	Normal approx.	True	Normal approx.	True	Normal approx.	True	Normal approx.	a_1	r
30	17	0·2566	0·2574	0·2194	0·2212							17	30
	18	·1376	·1388	·0981	·1002							18	
	19	·0630	·0643	·0348	·0365							19	
	20	·0243	·0253	·0096	·0106							20	
	21	·0078	·0085	·0020	·0024							21	
	22	·0021	·0024									22	
20	12	0·2269	0·2278	0·2060	0·2076	0·1715	0·1745					12	20
	13	·1053	·1068	·0852	·0873	·0564	·0592					13	
	14	·0392	·0408	·0270	·0287	·0128	·0144					14	
	15	·0114	·0126	·0064	·0073	·0019	·0025					15	
	16	·0025	·0031	·0011	·0014							16	
15	9	0·2884	0·2887	0·2760	0·2772	0·2572	0·2595	0·2330	0·2364			9	15
	10	·1312	·1325	·1163	·1185	·0954	·0985	·0715	·0755			10	
	11	·0453	·0473	·0358	·0380	·0242	·0265	·0134	·0156			11	
	12	·0113	·0129	·0077	·0090	·0040	·0049	·0014	·0020			12	
	13	·0019	·0027	·0011	·0016							13	
10	7	0·1589	0·1599	0·1495	0·1514	0·1367	0·1397	0·1226	0·1266	0·0894	0·0955	7	10
	8	·0458	·0486	·0399	·0429	·0324	·0357	·0251	·0285	·0115	·0147	8	
	9	·0078	·0101	·0061	·0081	·0042	·0058	·0026	·0038	·0005	·0011	9	
	10	·0006	·0014	·0004	·0010							10	
7	5	0·2179	0·2177	0·2119	0·2126	0·2038	0·2056	0·1950	0·1980	0·1749	0·1804	5	7
	6	·0558	·0594	·0514	·0553	·0458	·0501	·0401	·0448	·0286	·0338	6	
	7	·0062	·0096	·0053	·0084	·0042	·0068	·0032	·0055	·0015	·0031	7	
5	4	0·1810	0·1806	0·1766	0·1771	0·1709	0·1735	0·1648	0·1677	0·1517	0·1571	4	5
	5	·0281	·0339	·0261	·0320	·0236	·0295	·0211	·0270	·0163	·0220	5	

Table 9. *Case of unequal partition. Chances that $a \leqslant a_1$ and $a \geqslant a_1$*

(A) $m = \frac{3}{5}N$, $n = \frac{2}{5}N$

Partition			$m=60$, $n=40$		$m=36$, $n=24$		$m=24$, $n=16$		$m=18$, $n=12$		$m=12$, $n=8$				
r	Chance that	a_1	True	Normal approx.	True	Normal approx.	True	Normal approx.	True	Normal approx.	True	Normal approx.	a_1	Chance that	r
30	$a \leqslant a_1$	11	0·0020	0·0019									11	$a \leqslant a_1$	30
		12	·0074	·0074	0·0016	0·0020							12		
		13	·0230	·0231	·0084	·0093	—						13		
		14	·0601	·0604	·0320	·0337	0·0023	0·0050					14		
		15	·1330	·1339	·0936	·0957	·0270	·0329					15		
		16	·2512	·2531	·2148	·2165	·1311	·1348					16		
	$a \geqslant a_1$	20	0·2533	0·2531	0·2148	0·2165	0·1322	0·1348					20	$a \geqslant a_1$	
		21	·1323	·1339	·0936	·0957	·0318	·0329					21		
		22	·0580	·0604	·0320	·0337	·0045	·0050					22		
		23	·0209	·0231	·0084	·0093							23		
		24	·0061	·0074	·0016	·0020							24		
		25	·0014	·0019									25		
20	$a \leqslant a_1$	6	0·0027	0·0026	0·0010	0·0012							6	$a \leqslant a_1$	20
		7	·0114	·0112	·0060	·0063	0·0015	0·0021	—				7		
		8	·0381	·0378	·0255	·0262	·0112	·0128	0·0015	0·0033			8		
		9	·1019	·1021	·0816	·0829	·0526	·0555	·0290	·0260			9		
		10	·2211	·2232	·2005	·2028	·1665	·1695	·1170	·1218			10		
	$a \geqslant a_1$	14	0·2236	0·2232	0·2017	0·2028	0·1665	0·1695	0·1182	0·1218			14	$a \geqslant a_1$	
		15	·0994	·1021	·0798	·0829	·0526	·0555	·0241	·0260			15		
		16	·0341	·0378	·0233	·0262	·0112	·0128	·0026	·0033			16		
		17	·0086	·0112	·0048	·0063	·0015	·0021					17		
		18	·0015	·0026	·0006	·0012							18		
15	$a \leqslant a_1$	4	0·0053	0·0053	0·0032	0·0033	0·0013	0·0015					4	$a \leqslant a_1$	15
		5	·0236	·0233	·0171	·0173	·0098	·0106	0·0038	0·0052			5		
		6	·0776	·0775	·0650	·0657	·0481	·0499	·0301	·0335	—		6		
		7	·1948	·1968	·1804	·1827	·1588	·1618	·1317	·1358	0·0511	0·0616	7		
	$a \geqslant a_1$	11	0·1970	0·1968	0·1814	0·1827	0·1587	0·1618	0·1317	0·1358	0·0578	0·0616	11	$a \geqslant a_1$	
		12	·0734	·0775	·0614	·0657	·0458	·0499	·0301	·0335	·0036	·0051	12		
		13	·0188	·0233	·0138	·0173	·0082	·0106	·0038	·0052	—		13		
		14	·0029	·0053	·0018	·0033	·0008	·0015					14		
10	$a \leqslant a_1$	2	0·0088	0·0089	0·0067	0·0071	0·0045	0·0050	0·0026	0·0033	0·0004	0·0009	2	$a \leqslant a_1$	10
		3	·0457	·0453	·0395	·0398	·0318	·0329	·0241	·0260	·0099	·0131	3		
		4	·1538	·1549	·1447	·1464	·1322	·1348	·1182	·1218	·0849	·0910	4		
	$a \geqslant a_1$	8	0·1539	0·1549	0·1442	0·1464	0·1311	0·1348	0·1170	0·1218	0·0849	0·0910	8	$a \geqslant a_1$	
		9	·0386	·0453	·0334	·0398	·0270	·0329	·0209	·0260	·0099	·0131	9		
		10	·0044	·0089	·0034	·0071	·0023	·0050	·0015	·0033	·0004	·0009	10		
7	$a \leqslant a_1$	0	0·0012	0·0022	0·0009	0·0013	0·0006	0·0010	0·0004	0·0007			0	$a \leqslant a_1$	7
		1	·0156	·0189	·0134	·0140	·0109	·0118	·0086	·0097	0·0044	0·0059	1		
		2	·0884	·0956	·0827	·0832	·0756	·0770	·0681	·0704	·0521	·0564	2		
	$a \geqslant a_1$	6	0·1492	0·1587	0·1426	0·1450	0·1341	0·1378	0·1250	0·1300	0·1056	0·1127	6	$a \geqslant a_1$	
		7	·0241	·0385	·0216	·0306	·0186	·0269	·0156	·0232	·0102	0·0160	7		
5	$a \leqslant a_1$	0	0·0088	0·0099	0·0078	0·0090	0·0066	0·0080	0·0056	0·0070	0·0036	0·0051	0	$a \leqslant a_1$	5
		1	·0816	·0811	·0778	·0781	·0730	·0742	·0681	·0701	·0578	·0616	1		
	$a \geqslant a_1$	5	0·0725	0·0811	0·0690	0·0781	0·0646	0·0742	0·0601	0·0701	0·0511	0·0616	5	$a \geqslant a_1$	

Table 9 (*continued*)

(B) $m = \frac{4}{5}N,\ n = \frac{1}{5}N$

Partition			$m = 80,\ n = 20$		$m = 48,\ n = 12$		$m = 32,\ n = 8$	
r	Chance that	a_1	True	Normal approx.	True	Normal approx.	True	Normal approx.
30	$a \leqslant a_1$	18	0·0018	0·0014				
		19	·0084	·0073	0·0013	0·0020		
		20	·0306	·0288	·0106	·0125		
		21	·0884	·0874	·0521	·0548		
		22	·2046	·2078	·1667	·1685		
	$a \geqslant a_1$	26	0·2092	0·2078	0·1667	0·1685		
		27	·0824	·0874	·0521	·0548		
		28	·0227	·0288	·0106	·0125		
		29	·0039	·0073	·0013	·0020		
		30	·0003	·0014				
20	$a \leqslant a_1$	11	0·0040	0·0026	0·0013	0·0011		
		12	·0182	·0148	·0095	·0087	0·0016	0·0031
		13	·0638	·0600	·0460	·0448	·0218	·0255
		14	·1729	·1755	·1523	·1542	·1176	·1208
	$a \geqslant a_1$	18	0·1758	0·1755	0·1522	0·1542	0·1176	0·1208
		19	·0499	·0600	·0371	·0448	·0218	·0255
		20	·0066	·0148	·0041	·0087	·0016	·0031
15	$a \leqslant a_1$	7	0·0018	0·0009	0·0008	0·0004		
		8	·0107	·0074	·0064	·0049	0·0022	0·0024
		9	·0462	·0408	·0355	·0323	·0217	·0219
		10	·1470	·1480	·1329	·1338	·1115	·1133
	$a \geqslant a_1$	14	0·1453	0·1480	0·1294	0·1338	0·1079	0·1133
		15	·0262	·0408	·0206	·0323	·0141	·0219
10	$a \leqslant a_1$	4	0·0039	0·0019	0·0026	0·0013	0·0012	0·0008
		5	·0254	·0191	·0206	·0159	·0145	·0121
		6	·1095	·1068	·1012	·0988	·0893	·0882
	$a \geqslant a_1$	10	0·0951	0·1068	0·0868	0·0988	0·0761	0·0882
7	$a \leqslant a_1$	2	0·0033	0·0013	0·0024	0·0010	0·0015	0·0007
		3	·0282	·0203	·0246	·0181	·0201	·0155
		4	·1408	·1417	·1354	·1364	·1281	·1293
	$a \geqslant a_1$	7	0·1985	0·1910	0·1906	0·1848	0·1805	0·1776
5	$a \leqslant a_1$	1	0·0053	0·0022	0·0045	0·0021	0·0035	0·0016
		2	·0531	·0434	·0499	·0430	·0457	·0383
	$a \geqslant a_1$	5	0·3193	0·2841	0·3135	0·2835	0·3060	0·2776

(C) $m = \frac{9}{10}N,\ n = \frac{1}{10}N$

Partition			$m = 90,\ n = 10$	
r	Chance that	a_1	True	Normal approx.
30	$a \leqslant a_1$	22	0·0009	0·0006
		23	·0073	·0057
		24	·0388	·0352
		25	·1384	·1388
	$a \geqslant a_1$	29	0·1356	0·1388
		30	·0229	·0352
20	$a \leqslant a_1$	14	0·0039	0·0019
		15	·0254	·0191
		16	·1095	·1068
	$a \geqslant a_1$	20	0·0951	0·1068
15	$a \leqslant a_1$	9	0·0006	0·0001
		10	·0063	·0027
		11	·0408	·0316
		12	·1705	·1765
	$a \geqslant a_1$	15	0·1808	0·1765
10	$a \leqslant a_1$	5	0·0006	0·0001
		6	·0082	·0029
		7	·0600	·0486
		8	·2615	·2902
	$a \geqslant a_1$	10	0·3305	0·2902
7	$a \leqslant a_1$	3	0·0016	0·0003
		4	·0207	·0096
		5	·1442	·1492
	$a \geqslant a_1$	7	0·4667	0·3974
5	$a \leqslant a_1$	2	0·0067	0·0006
		3	·0769	·0538
	$a \geqslant a_1$	5	0·4163	0·5000

2×2 TABLES. A NOTE ON E. S. PEARSON'S PAPER

By G. A. BARNARD

As Prof. Pearson has kindly shown me the proof of his paper, I should like to make the following further remarks.

1. If we have a sample of N from a population in which there is a chance p that an individual will have a character A, we can represent it in the form

$$x_1, x_2, \ldots, x_i, \ldots, x_N,$$

where x_i is 1 or 0 according as to whether the ith member has A or not.* Regarding the x's as quantitative variables, we have by classical results the unbiased estimates

$$\hat{p} = x. = (\Sigma x_i)/N \quad \text{and} \quad \hat{\sigma}^2 = (\Sigma(x_i - x.)^2)/(N-1).$$

If r of the x's are 1, while s are 0, we find

$$\hat{p} = r/N \quad \text{and} \quad \hat{\sigma}^2 = rs/N(N-1).$$

Using this unbiased estimate of variance in Prof. Pearson's para. 43, we get, instead of his (28·2),

$$\frac{d}{s_d} = \frac{a - rm/N}{\sqrt{\dfrac{mnrs}{N^2(N-1)}}}, \tag{1}$$

agreeing exactly with his (22).

2. To carry the argument further, in classical theory, if we have two samples

$$(x_1, x_2, \ldots, x_i, \ldots, x_m) \quad \text{and} \quad (y_1, y_2, \ldots, y_j, \ldots, y_n)$$

to test whether the samples come from the same normal population we take

$$t = \frac{x. - y.}{s}\sqrt{\frac{mn}{m+n}},$$

where $x. = (\Sigma x_i)/m$, $y. = (\Sigma y_j)/n$, and

$$s^2 = \frac{\Sigma(x_i - x.)^2 + \Sigma(y_j - y.)^2}{m+n-2}, \tag{2}$$

and use tables of the t distribution for $(m+n-2)$ degrees of freedom.

It is common practice to neglect departures from normality in applying this test. If we do so, and apply it to our qualitative case along the lines indicated above, we get

$$t = \frac{a - rm/N}{\sqrt{\dfrac{acn + bdm}{N(N-2)}}},$$

which, if we are justified in our neglect of departures from normality, should be distributed as t on $(N-2)$ degrees of freedom.

* For a similar argument see B. L. Welch (1938, p. 155).

3. To obtain the formula (1) on these lines, we have in effect to commit the well-known fallacy of replacing s^2 as given by (2), by

$$s'^2 = \frac{\Sigma(x_i - m')^2 + \Sigma(y_j - m')^2}{m+n-1}, \tag{3}$$

where $$m' = (\Sigma x_i + \Sigma y_j)/(m+n).$$

We are led to ask why (3) should be approximately correct (and in fact it is better than (2)) in the qualitative case, while (2) is preferred in the quantitative case.

4. The simplest reason for preferring (2) to (3) in the quantitative case is that s'^2 is not independent of $(x. - y.)$, so that the conditions for validity of the t distribution are not satisfied. In our qualitative case this argument loses validity, since neither s^2 nor s'^2 is independent of $(x. - y.)$.

The second reason for preferring (2) to (3) in the quantitative case is more complicated, but for our purposes it reduces essentially to the fact that, in the case of normal distributions, and *only in this case*, the mean and variance of samples are independently distributed, so that the common mean value of the populations, estimated by m', is irrelevant to the test for differences. In our qualitative case, on the other hand, m' contributes to our knowledge of the variance.

5. If we apply Pitman's 'absolute' analogue of the t test to our case, we arrive at the hypergeometric series of Prof. Pearson's Problem I. But Bartlett's argument, showing the convergence of Pitman's test and the t test, will apply here only in very large samples, because of the finite probability of obtaining observed values which coincide.

6. From the above point of view, Prof. Pearson's analysis of his Problem II may be regarded in one sense as an examination of the effect of large departures from normality on the t test. In this light, his conclusions given in paras. 51 and 52 are seen to extend to the t test, as well as to the 2×2 table problem.

7. If I may state my personal attitude, it is that statistics is a branch of applied mathematics, like symbolic logic or hydrodynamics. Examination of foundations is desirable, but it must be remembered that undue emphasis on niceties is a disease to which persons with mathematical training are specially prone. In pure mathematics itself there are disputes on foundations which closely parallel the disputes over the foundations of statistics. The lesson to be drawn is, that while statistics is a most valuable aid to judgement, it cannot wholly replace it.

8. Finally, it must be emphasized that the order of printing of Prof. Pearson's paper and my own reflects Prof. Pearson's generosity rather than the historical order of events. Much of his paper was, unknown to me, given in lectures before the war; whereas my work on the problem began only in 1943. Since then I have owed much both to Prof. Pearson's published work and to discussions which I have been privileged to have with him.

REFERENCE

WELCH, B. L. (1938). *Biometrika*, **30**, 155.

THE CUMULANTS OF THE Z AND OF THE LOGARITHMIC χ^2 AND t DISTRIBUTIONS

By JOHN WISHART

School of Agriculture, Cambridge

Explicit expressions for the exact cumulants of Fisher's z-distribution do not appear ever to have been published. They were therefore worked out, and appear in § 2 of this paper. It afterwards appeared that the logical method of presentation was to deal with the similar problem for $\frac{1}{2}\log(\chi^2/n)$,* since the z-distribution involves the simple difference of two such functions which are independent. This led to § 1. Since writing this paper, Bartlett & Kendall (1946) have published the same result in the form of the cumulants of $\log s^2$, and have given graphical and tabular representations for varying n up to 20. The solution is, of course, implicit in Cornish & Fisher's (1937) statement of the moment generating function, while Mr C. R. Rao has informed me that he reached the same result in work done for an M.A. Thesis of the University of Calcutta (unpublished). § 1 has accordingly been shortened, but is retained in view of the additional formulae to those of Bartlett and Kendall.

1. The logarithmic χ^2 distribution

The distribution of χ^2, for n degrees of freedom, is given by

$$\frac{1}{\Gamma(\frac{1}{2}n)}(\tfrac{1}{2}\chi^2)^{\frac{1}{2}n-1}e^{-\frac{1}{2}\chi^2}d(\tfrac{1}{2}\chi^2).$$

As pointed out by Cornish & Fisher (1937), the mean value of $\exp\{\frac{1}{2}it\log(\chi^2/n)\}$

i.e. of $\exp\{\frac{1}{2}it\log(\frac{1}{2}\chi^2)-\frac{1}{2}it\log(\frac{1}{2}n)\}$

is the moment generating function of the distribution of $\frac{1}{2}\log(\chi^2/n)$, namely

$$M=\frac{\Gamma\frac{1}{2}(n+it)}{\Gamma(\frac{1}{2}n)}\exp\{-\tfrac{1}{2}it\log(\tfrac{1}{2}n)\}.$$

The cumulant generating function is

$$K=\log M=-\tfrac{1}{2}it\log(\tfrac{1}{2}n)+\log\Gamma\tfrac{1}{2}(n+it)-\log\Gamma(\tfrac{1}{2}n).$$

The cumulants of the distribution of $\frac{1}{2}\log(\chi^2/n)$ are readily written down by differentiating K successively with respect to it and at each stage putting $t=0$. We have in fact

$$\begin{aligned}\kappa_1&=-\tfrac{1}{2}\log(\tfrac{1}{2}n)+\frac{d}{dn}\log\Gamma(\tfrac{1}{2}n)\\&=-\frac{1}{2}\left\{\log a+\mathrm{Lt}_{s\to 1}\left(\zeta(s,a)-\frac{1}{s-1}\right)\right\}\end{aligned}\qquad(1)$$

and $$\kappa_s=\frac{(-1)^s(s-1)!}{2^s}\zeta(s,a)\quad(s>1,\ a=\tfrac{1}{2}n),$$

where $\zeta(s,a)$ denotes the generalized Zeta-function

$$\zeta(s,a)=\sum_{j=0}^{\infty}\frac{1}{(a+j)^s}.$$

* All logarithms in this paper are to base e.

The cumulants may be readily computed by throwing them into the form

$$2\kappa_1 = \psi(\tfrac{1}{2}n) - \log(\tfrac{1}{2}n),$$

$$2^s\kappa_s = \psi^{(s-1)}(\tfrac{1}{2}n), \tag{2}$$

where $\psi(x) = d\{\log\Gamma(x)\}/dx, \quad \psi^{(s-1)}(x) = d^s\{\log\Gamma(x)\}/dx^s.$

$\psi(x)$ is variously called the Psi or Digamma function, and its derivatives have been called the Trigamma, Tetragamma, etc. Functions, and the series the Polygamma Functions. These functions have been computed in some considerable detail. For n up to 22 the mean and variance can be got from Elinor Pairman's 'Tables of the Digamma and Trigamma functions' (1919). Tables up to Pentagamma appear in Vol. I of the British Association's *Mathematical Tables* (1931), but with certain gaps which, although intended to be bridged by reduction formulae, render the tables less generally useful (for n less than 22) than H. T. Davis's Tables (1933, 1935). Table 10 of Vol. I gives all that is required for $\psi(x)$; in Vol. II, Tables 14–16, 18–20, 22–24 and 26–28 cover a wide range up to Hexagamma.

As shown by Bartlett & Kendall (1946), the approach to normality is very slow. For $n = 24$ (the limit for n_1 of the z table of Fisher & Yates (1943), which provides percentage points for the distribution under consideration in the line $n_2 = \infty$) the cumulants have been worked out to κ_6, the last being specially computed from its formula given below. The gamma ratios are $\gamma_1 = -0{\cdot}295$, $\gamma_2 = 0{\cdot}174$, $\gamma_3 = -0{\cdot}154$ and $\gamma_4 = 0{\cdot}175$, and $|\gamma|$ increases thereafter at this level of n instead of tending to zero. Approximate percentage points may, however, be worked out by using the formulae at the foot of the z table, putting $n_2 = \infty$.

For small n, we note that

$$\zeta(s) = \sum_{j=1}^{\infty}\frac{1}{j^s} = \frac{1}{1^s}+\frac{1}{2^s}+\frac{1}{3^s}+\dots+\frac{1}{(r-1)^s}+\zeta(s,r) \quad r \text{ an integer},$$

$$\zeta(s).(1-2^{-s}) = \frac{1}{1^s}+\frac{1}{3^s}+\frac{1}{5^s}+\dots+\infty$$

$$= \frac{1}{1^s}+\frac{1}{3^s}+\frac{1}{5^s}+\dots+\frac{1}{(2r-1)^s}+2^{-s}\zeta(s,r+\tfrac{1}{2}).$$

We thus get, *for* $n = 2r$,

$$\kappa_s = \frac{(-1)^s(s-1)!}{2^s}\left\{\zeta(s)-\left(1+\frac{1}{2^s}+\frac{1}{3^s}+\dots+\frac{1}{(\frac{1}{2}n-1)^s}\right)\right\} \quad (s>1), \tag{3}$$

in which the terms in $\{\dots\}$ reduce to $\zeta(s)$ for $n = 2$.

For $n = 2r+1$

$$\kappa_s = (-1)^s(s-1)!\left\{\zeta(s)(1-2^{-s})-\left(1+\frac{1}{3^s}+\frac{1}{5^s}+\dots+\frac{1}{(n-2)^s}\right)\right\} \quad (s>1), \tag{4}$$

in which the terms in $\{\dots\}$ reduce to $\zeta(s)(1-2^{-s})$ for $n = 1$.

In the special case of $s = 1$ we have

$$n = 2r \qquad \kappa_1 = -\tfrac{1}{2}(\gamma+\log\tfrac{1}{2}n)+\frac{1}{2}\left(1+\frac{1}{2}+\frac{1}{3}+\dots+\frac{1}{\frac{1}{2}n-1}\right),$$

$$n = 2r+1 \qquad \kappa_1 = -\tfrac{1}{2}(\gamma+\log 2n)+\left(1+\frac{1}{3}+\frac{1}{5}+\dots+\frac{1}{n-2}\right). \tag{5}$$

For $n = 2$ and 1 respectively these expressions reduce to the first bracket. $\zeta(s)$ can be got from tables, and in particular

$$\zeta(2m) = 2^{2m-1}\pi^{2m}B_m/(2m)!,$$

where the B's are the Bernoulli numbers $B_1 = \frac{1}{6}$, $B_2 = \frac{1}{30}$, $B_3 = \frac{1}{42}$, $B_4 = \frac{1}{30}$, $B_5 = \frac{5}{66}$, etc. γ is Euler's constant. For reference we may quote:

$$\gamma = 0{\cdot}57721\ 56649, \qquad \zeta(4) = 1{\cdot}08232\ 32337,$$
$$\zeta(2) = 1{\cdot}64493\ 40668, \qquad \zeta(5) = 1{\cdot}03692\ 77551,$$
$$\zeta(3) = 1{\cdot}20205\ 69032, \qquad \zeta(6) = 1{\cdot}01734\ 30620.$$

Note that
$$\mathrm{Lt}_{s\to 1}\left(\zeta(s,a) - \frac{1}{s-1}\right) = \gamma + \sum_{j=0}^{\infty}\left(\frac{1}{a+j} - \frac{1}{1+j}\right), \quad (R(a) > 0).$$

For large n, asymptotic formulae for the Zeta-function may be used, and we get

$$\kappa_1 \sim -\frac{1}{2n} - \sum_{j=1}^{\infty}\frac{(-4)^{j-1}B_j}{jn^{2j}}$$
$$\sim -\frac{1}{2n} - \frac{1}{6n^2} + \frac{1}{15n^4} - \frac{8}{63n^6} + \frac{8}{15n^8} - \frac{64}{33n^{10}} + \ldots, \tag{6}$$
$$\kappa_s \sim (-1)^s\left[\frac{(s-2)!}{2n^{s-1}} + \frac{(s-1)!}{2n^s} + \frac{2}{n^s}\sum_{j=1}^{\infty}\frac{(-4)^{j-1}B_j(2j+s-2)!}{(2j)!\,n^{2j-1}}\right] \quad (s > 1).$$

We may note in passing that not only may this general expression for κ_s be applied to the special case $s = 1$ with the proviso that the first term in that case is dropped, but also that κ_2 may be obtained from $\kappa_1 + \frac{1}{2}\log(\frac{1}{2}n)$ by term-by-term differentiation with respect to n, and likewise κ_3 from κ_2, κ_4 from κ_3, etc., by similar term-by-term differentiation. This follows from a property of the Zeta-function. It is therefore not necessary to write down the explicit expressions for κ_2, κ_3, etc., but we may note that their leading terms are $\frac{1}{2}n^{-1}$, $-\frac{1}{2}n^{-2}$, n^{-3}, $-3n^{-4}$, $12n^{-5}$, etc., so that the leading terms of γ_1 and γ_2 are $-\sqrt{(2/n)}$ and $4/n$ respectively, while γ_r is $O(n^{-\frac{1}{2}r})$. More exactly we have, writing $n' = n-1$,

$$\kappa_2 = \frac{1}{2n'}\left(1 - \frac{1}{3n'^2} + \frac{7}{15n'^4} + O\left(\frac{1}{n'^6}\right)\right)$$

with corresponding expressions for κ_3, κ_4, etc., obtained by differentiation with respect to n', and

$$\gamma_1 = -\sqrt{\frac{2}{n'}}\left(1 - \frac{1}{2n'^2} + O\left(\frac{1}{n'^4}\right)\right), \qquad \gamma_r \sim (-1)^r r!\left(\frac{2}{n'}\right)^{\frac{1}{2}r},$$
$$\gamma_2 = \frac{4}{n'}\left(1 - \frac{4}{3n'^2} + O\left(\frac{1}{n'^4}\right)\right), \qquad \frac{\gamma_r}{\gamma_{r-1}} \sim -r\sqrt{\left(\frac{2}{n'}\right)}.$$

Finally, if instead of the distribution of $\frac{1}{2}\log(\chi^2/n)$ we are interested in the distribution of $\log(s^2)$, where s^2 is an estimate of σ^2 based on n degrees of freedom, we have

$$\log(\chi^2/n) = \log s^2 - \log\sigma^2,$$

and thus for the distribution of $\log(s^2)$ we have

$$\kappa_1 = \log\left(\frac{2\sigma^2}{n}\right) + 2\frac{d}{dn}\log\Gamma(\tfrac{1}{2}n)$$
$$= \log\left(\frac{2\sigma^2}{n}\right) - \mathrm{Lt}_{s\to 1}\left\{\zeta(s,a) - \frac{1}{s-1}\right\}$$

and
$$\kappa_s = (-1)^s(s-1)!\,\zeta(s,a) \quad (s > 1, a = \tfrac{1}{2}n),$$

while the γ ratios are the same as for $\frac{1}{2}\log(\chi^2/n)$. Obviously $\log\chi$ and $\log s$ can be treated similarly. See Bartlett & Kendall (1946).

2. THE z DISTRIBUTION

The distribution of $z = \frac{1}{2}\log(s_1^2/s_2^2)$, where s_1^2 and s_2^2 are independent estimates of a variance σ^2, based respectively on ν_1 and ν_2 degrees of freedom, is obviously that of

$$\tfrac{1}{2}\log(\chi_1^2/\nu_1)-\tfrac{1}{2}\log(\chi_2^2/\nu_2)$$

and its cumulants may therefore be at once derived from those of the logarithmic χ^2 distribution. The cumulant generating function is

$$K = \log M = \tfrac{1}{2}it\log(\nu_2/\nu_1)+\log\Gamma\tfrac{1}{2}(\nu_1+it)+\log\Gamma\tfrac{1}{2}(\nu_2-it)-\log\Gamma(\tfrac{1}{2}\nu_1)-\log\Gamma(\tfrac{1}{2}\nu_2).$$

Further, we have

$$\begin{aligned}\kappa_1 &= \tfrac{1}{2}\log\frac{\nu_2}{\nu_1}+\frac{d}{d\nu_1}\log\Gamma(\tfrac{1}{2}\nu_1)-\frac{d}{d\nu_2}\log\Gamma(\tfrac{1}{2}\nu_2)\\ &= \frac{1}{2}\left\{\log\frac{\nu_2}{\nu_1}+\mathrm{Lt}_{s\to 1}\left(\zeta(s,a_2)-\zeta(s,a_1)\right)\right\} \qquad (7)\\ \kappa_s &= 2^{-s}(s-1)!\{\zeta(s,a_2)+(-1)^s\zeta(s,a_1)\} \quad (s>1,\ a_1=\tfrac{1}{2}\nu_1,\ a_2=\tfrac{1}{2}\nu_2).\end{aligned}$$

For computing purposes these may be thrown into the forms

$$\begin{aligned}2\kappa_1 &= \log(\nu_2/\nu_1)+\psi(\tfrac{1}{2}\nu_1)-\psi(\tfrac{1}{2}\nu_2),\\ 2^s\kappa_s &= \psi^{(s-1)}(\tfrac{1}{2}\nu_1)+(-1)^s\psi^{(s-1)}(\tfrac{1}{2}\nu_2) \quad (s>1). \qquad (8)\end{aligned}$$

To illustrate, let us take $\nu_1 = 24$, $\nu_2 = 60$. We then have from the Polygamma tables (except for κ_6, which was specially computed):

$$\kappa_1 = -0{\cdot}0127\,429, \quad \kappa_3 = -0{\cdot}0007\,998, \quad \kappa_5 = -0{\cdot}0000\,104,$$
$$\kappa_2 = 0{\cdot}0301\,992, \quad \kappa_4 = 0{\cdot}0000\,867, \quad \kappa_6 = 0{\cdot}0000\,019,$$
$$\sigma = \sqrt{\kappa_2} = 0{\cdot}1737\,792,$$

$\gamma_1 = -0{\cdot}152$, $\gamma_2 = 0{\cdot}095$ (or $\beta_1 = 0{\cdot}023$, $\beta_2 = 3{\cdot}095$), indicating the degree and nature of the departure from normality. γ_3 and γ_4 are $-0{\cdot}066$ and $0{\cdot}067$ respectively.

If as a first approximation we assume that for ν_1 and ν_2 of the order of the numbers chosen in this example, or higher, z is distributed normally with mean and variance given by the above formulae, we obtain approximate percentage points, e.g. for the 95 and 5 % points we can subtract and add $1{\cdot}6449\sigma$ from and to the mean. The result in the present case is to give us $-0{\cdot}299$ and $0{\cdot}273$, the correct values being $-0{\cdot}306$ and $0{\cdot}265$. The approximation is adequate to almost two figure accuracy, and is evidently useful when we only require to know whether an observed z is significant or not. A better approximation is provided by the formulae attached to the z tables (see Fisher & Yates (1943)), which yield $-0{\cdot}3045$ and $0{\cdot}2653$ as against the correct values of $-0{\cdot}3055$ (see Thompson (1941)) and $0{\cdot}2654$.

Explicit algebraic expressions are readily written down for the cumulants for small ν_1 and ν_2, using the same method as for the logarithmic χ^2 distribution. Where it is necessary to do so, ν_1 will be assumed less than ν_2. In the contrary case we need only interchange ν_1 and ν_2, changing the sign of the odd cumulants in so doing. The odd cumulants are zero when $\nu_1 = \nu_2$. We have

Even cumulants $\qquad (r = 2s,\ s>0)$

$\nu_1 = 2p$, $\nu_2 = 2q$

$$\kappa_r = 2(r-1)!\left[\zeta(r)\,2^{-r}-\left(\frac{1}{2^r}+\frac{1}{4^r}+\dots+\frac{1}{(\nu_1-2)^r}\right)-\frac{1}{2}\left(\frac{1}{\nu_1^r}+\frac{1}{(\nu_1+2)^r}+\dots+\frac{1}{(\nu_2-2)^r}\right)\right], \qquad (9)$$

$\nu_1 = 2p+1,\ \nu_2 = 2q+1$

$$\kappa_r = 2(r-1)!\left[\zeta(r)(1-2^{-r}) - \left(\frac{1}{1^r}+\frac{1}{3^r}+\dots+\frac{1}{(\nu_1-2)^r}\right) - \frac{1}{2}\left(\frac{1}{\nu_1^r}+\frac{1}{(\nu_1+2)^r}+\dots+\frac{1}{(\nu_2-2)^r}\right)\right]. \quad (10)$$

$\nu_1 = \nu_2$. Drop out the last bracket of terms in the above two cases.

$\nu_1 = 2p,\ \nu_2 = 2q+1$

$$\kappa_r = (r-1)!\left[\zeta(r) - \left(\frac{1}{2^r}+\frac{1}{4^r}+\dots+\frac{1}{(\nu_1-2)^r}\right) - \left(1+\frac{1}{3^r}+\frac{1}{5^r}+\dots+\frac{1}{(\nu_2-2)^r}\right)\right]. \quad (11)$$

$\nu_1 = 2p+1,\ \nu_2 = 2q$. Interchange ν_1 and ν_2 in this last case.

Odd cumulants $\qquad (r = 2s+1,\ s>0)$

$\nu_1 = 2p,\ \nu_2 = 2q$, or $\nu_1 = 2p+1,\ \nu_2 = 2q+1$

$$\kappa_r = -(r-1)!\left[\frac{1}{\nu_1^r}+\frac{1}{(\nu_1+2)^r}+\dots+\frac{1}{(\nu_2-2)^r}\right]. \quad (12)$$

$\nu_1 = 2p,\ \nu_2 = 2q+1$

$$\kappa_r = (r-1)!\left[\zeta(r)(1-2^{1-r}) + \left(\frac{1}{2^r}+\frac{1}{4^r}+\dots+\frac{1}{(\nu_1-2)^r}\right) - \left(1+\frac{1}{3^r}+\frac{1}{5^r}+\dots+\frac{1}{(\nu_2-2)^r}\right)\right]. \quad (13)$$

$\nu_1 = 2p+1,\ \nu_2 = 2q$. Interchange ν_1 and ν_2 in this last case, and change the sign of κ_r.

In the special case of $s = 1$, we have

$\nu_1 = 2p,\ \nu_2 = 2q$, or $\nu_1 = 2p+1,\ \nu_2 = 2q+1$

$$\kappa_1 = \tfrac{1}{2}\log\left(\frac{\nu_2}{\nu_1}\right) - \left(\frac{1}{\nu_1}+\frac{1}{\nu_1+2}+\dots+\frac{1}{\nu_2-2}\right). \quad (14)$$

$\nu_1 = 2p,\ \nu_2 = 2q+1$

$$\kappa_1 = \tfrac{1}{2}\log\left(\frac{\nu_2}{\nu_1}\right) + \log 2 + \left(\frac{1}{2}+\frac{1}{4}+\dots+\frac{1}{\nu_1-2}\right) - \left(1+\frac{1}{3}+\frac{1}{5}+\dots+\frac{1}{\nu_2-2}\right). \quad (15)$$

$\nu_1 = 2p+1,\ \nu_2 = 2q$. Interchange ν_1 and ν_2 in this last case and change the sign of κ_1.

For large ν_1 *and* ν_2, a combination of the asymptotic formulae already given readily yields the following results:

$$\kappa_1 \sim \frac{1}{2}\left(\frac{1}{\nu_2}-\frac{1}{\nu_1}\right) + \sum_{j=1}^{\infty}\frac{(-4)^{j-1}B_j}{j}\left(\frac{1}{\nu_2^{2j}}-\frac{1}{\nu_1^{2j}}\right),$$

the numerical coefficients being as for the κ_1 of $\frac{1}{2}\log(\chi^2/n)$,

$$\kappa_s \sim \frac{(s-2)!}{2}\left(\frac{1}{\nu_2^{s-1}}+\frac{(-1)^s}{\nu_1^{s-1}}\right) + \frac{(s-1)!}{2}\left(\frac{1}{\nu_2^s}+\frac{(-1)^s}{\nu_1^s}\right)$$
$$+2\sum_{j=1}^{\infty}\frac{(-4)^{j-1}B_j(2j+s-2)!}{(2j)!}\left(\frac{1}{\nu_2^{2j+s-1}}+\frac{(-1)^s}{\nu_1^{2j+s-1}}\right). \quad (16)$$

We may put $s = 1$ in κ_s provided we drop the first term. We note also that κ_2 and higher cumulants can be written down immediately by differentiating the terms in ν_2 and ν_1 of $\kappa_1 - \frac{1}{2}\log(\nu_2/\nu_1)$ successively with respect to $-\nu_2$ and ν_1 respectively.

These are the results given by Cornish & Fisher (1937), whose formulae can be extended at sight by means of the results of this paper. A first approximation not only gives the familiar results

$$\kappa_1 \sim \frac{1}{2}\left(\frac{1}{\nu_2}-\frac{1}{\nu_1}\right), \quad \kappa_2 \sim \frac{1}{2}\left(\frac{1}{\nu_2}+\frac{1}{\nu_1}\right),$$

but also the more general $\qquad \kappa_s \sim \dfrac{(s-2)!}{2}\left(\dfrac{1}{\nu_2^{s-1}}+\dfrac{(-1)^s}{\nu_1^{s-1}}\right),$

but it should be noted that for all $s > 1$ a second approximation, which takes in an additional term, is

$$\kappa_s \sim \frac{(s-2)!}{2}\left(\frac{1}{(\nu_2-1)^{s-1}}+\frac{(-1)^s}{(\nu_1-1)^{s-1}}\right).$$

The accuracy of the asymptotic approximation at the limits of the z table given by Fisher & Yates (1943) can be seen by applying it to our example ($\nu_1 = 24$, $\nu_2 = 60$). The numbers of terms which are significant in the eighth place (needed for final accuracy to 7 decimal places), are three for κ_1, four for κ_2 and κ_3, and three for κ_4, κ_5 and κ_6. The first term for κ_6, namely $12(\nu_2^{-5}+\nu_1^{-5})$, yields 0·0000 015, rather more than 20 % too low. To use

$$12\{(\nu_2-1)^{-5}+(\nu_1-1)^{-5}\}$$

would give 0·0000 019, about 2 % too high.

Should ν_1 or ν_2 be only moderate in size, the other being large, we may make use of the relation

$$\zeta(s,a) = \frac{1}{a^s}+\frac{1}{(a+1)^s}+\ldots+\frac{1}{(a+r-1)^s}+\zeta(s,a+r), \quad (r \text{ an integer}),$$

where a is one-half of the smaller of ν_1 or ν_2, to convert our formulae into forms in which asymptotic expansions may be applied to both of the Zeta-functions. We then have ($\nu_1 < \nu_2$):

$$\kappa_1 = \frac{1}{2}\left[\log\left(\frac{\nu_2}{\nu_1}\right)+\mathrm{Lt}_{s\to 1}\{\zeta(s,\tfrac{1}{2}\nu_2)-\zeta(s,\tfrac{1}{2}\nu_1+r)\}\right]-\left(\frac{1}{\nu_1}+\frac{1}{\nu_1+2}+\ldots+\frac{1}{\nu_1+2r-2}\right),$$

$$\kappa_s = (s-1)!\left[2^{-s}\{\zeta(s,\tfrac{1}{2}\nu_2)+(-1)^s\zeta(s,\tfrac{1}{2}\nu_1+r)\}+(-1)^s\left\{\frac{1}{\nu_1^s}+\frac{1}{(\nu_1+2)^s}+\ldots+\frac{1}{(\nu_1+2r-2)^s}\right\}\right], \tag{17}$$

and
$$\zeta(s,n) \sim \frac{1}{(s-1)\,n^{s-1}}+\frac{1}{2n^s}+\frac{1}{(s-1)!\,n^s}\sum_{j=1}^{\infty}\frac{(-1)^{j-1}B_j(2j+s-2)!}{(2j)!\,n^{2j-1}}.$$

Particular cases of some interest arise (i) when $r = \frac{1}{2}(\nu_2-\nu_1)$, ν_1 and ν_2 being either both odd or both even, and (ii) when $r = \frac{1}{2}(\nu_2-\nu_1+1)$, ν_1 (or ν_2) being even and ν_2 (or ν_1) odd. In the former case the first term within squared brackets in κ_s is $2^{-s}(1+(-1)^s)\,\zeta(s,\frac{1}{2}\nu_2)$, which is zero when s is odd and $2^{1-s}\zeta(s,\frac{1}{2}\nu_2)$ when s is even. In the latter we have

$$2^{-s}\{\zeta(s,\tfrac{1}{2}\nu_2)+(-1)^s\zeta(s,\tfrac{1}{2}(\nu_2+1))\}$$

which is $\zeta(s,\nu_2)$ when s is even. With s odd we are concerned with the difference of two Zeta-functions in which the a's differ by one-half, and the expression may be written

$$\sum_{j=0}^{\infty}\frac{(-1)^j}{(\nu_2+j)^s} = \frac{1}{(s-1)!}\left(\frac{d}{d\nu_2}\right)^{s-1}\sum_{j=0}^{\infty}\frac{(-1)^j}{\nu_2+j}$$

and
$$\begin{aligned}\sum_{j=0}^{\infty}\frac{(-1)^j}{\nu_2+j} &= \int_0^1\frac{x^{\nu_2-1}\,dx}{1+x}\\ &= \sum_{j=0}^{\infty}\frac{j!(\nu_2-1)!}{2^{j+1}(\nu_2+j)!} \quad \text{on integration by parts}\\ &= \frac{1}{2\nu_2}+\sum_{j=1}^{\infty}\frac{(-1)^{j-1}(2^{2j}-1)B_j}{2j\,\nu_2^{2j}}\end{aligned}$$

on expansion in powers of ν_2^{-1}. This asymptotic expansion is an interesting one in which the early coefficients are very simple, for the series is

$$\frac{1}{2\nu_2}+\frac{1}{4\nu_2^2}-\frac{1}{8\nu_2^4}+\frac{1}{4\nu_2^6}-\frac{17}{16\nu_2^8}+\ldots. \tag{18}$$

The various cases are set out below:

Even cumulants $(r = 2s,\ s > 0)$

$\nu_1 = 2p,\ \nu_2 = 2q$, or $\nu_1 = 2p+1,\ \nu_2 = 2q+1$

$$\kappa_r = (r-1)!\left\{\frac{1}{\nu_1^r}+\frac{1}{(\nu+2)^r}+\dots+\frac{1}{(\nu_2-2)^r}\right\}$$
$$+\frac{(r-2)!}{\nu_2^{r-1}}+\frac{(r-1)!}{\nu_2^r}-\frac{1}{\nu_2^r}\sum_{j=1}^{\infty}\frac{(-4)^j B_j(2j+r-2)!}{(2j)!\,\nu_2^{2j-1}}, \tag{19}$$

$\nu_1 = 2p,\ \nu_2 = 2q+1$

$$\kappa_r = (r-1)!\left\{\frac{1}{\nu_1^r}+\frac{1}{(\nu_1+2)^r}+\dots+\frac{1}{(\nu_2-1)^r}\right\}$$
$$+\frac{(r-2)!}{\nu_2^{r-1}}+\frac{(r-1)!}{2\nu_2^r}-\frac{1}{\nu_2^r}\sum_{j=1}^{\infty}\frac{(-1)^j B_j(2j+r-2)!}{(2j)!\,\nu_2^{2j-1}}. \tag{20}$$

Odd cumulants $(r = 2s+1,\ s > 0)$

$\nu_1 = 2p,\ \nu_2 = 2q$, or $\nu_1 = 2p+1,\ \nu_2 = 2q+1$

$$\kappa_r = -(r-1)!\left\{\frac{1}{\nu_1^r}+\frac{1}{(\nu_1+2)^r}+\dots+\frac{1}{(\nu_2-2)^r}\right\}, \tag{21}$$

$\nu_1 = 2p,\ \nu_2 = 2q+1$

$$\kappa_r = -(r-1)!\left\{\frac{1}{\nu_1^r}+\frac{1}{(\nu_1+2)^r}+\dots+\frac{1}{(\nu_2-1)^r}\right\}$$
$$+\frac{(r-1)!}{2\nu_2^r}+\frac{1}{\nu_2^r}\sum_{j=1}^{\infty}\frac{(-1)^{j-1}(2^{2j}-1)B_j(2j+r-2)!}{(2j)!\,\nu_2^{2j-1}}. \tag{22}$$

In the special case of $s = 1$, we have

$\nu_1 = 2p,\ \nu_2 = 2q$, or $\nu_1 = 2p+1,\ \nu_2 = 2q+1$

$$\kappa_1 = \tfrac{1}{2}\log\left(\frac{\nu_2}{\nu_1}\right)-\left(\frac{1}{\nu_1}+\frac{1}{\nu_1+2}+\dots+\frac{1}{\nu_2-2}\right),$$

$\nu_1 = 2p,\ \nu_2 = 2q+1$

$$\kappa_1 = \tfrac{1}{2}\log\left(\frac{\nu_2}{\nu_1}\right)-\left(\frac{1}{\nu_1}+\frac{1}{\nu_1+2}+\dots+\frac{1}{\nu_2-1}\right)+\frac{1}{2\nu_2}+\sum_{j=1}^{\infty}\frac{(-1)^{j-1}(2^{2j}-1)B_j}{2j\nu_2^{2j}}. \tag{23}$$

3. The logarithmic t-distribution

When $\nu_1 = 1$, $z = \log t$, and we thus have as a special case for the distribution of $\log t$ for $\nu_2 = n$ degrees of freedom:

$$2\kappa_1 = \log n + \mathrm{Lt}_{s\to 1}\{\zeta(s,\tfrac{1}{2}n)-\zeta(s,\tfrac{1}{2})\} \tag{24}$$
$$= \log n + \psi(\tfrac{1}{2})-\psi(\tfrac{1}{2}n) \qquad (\psi(\tfrac{1}{2}) = -\gamma-2\log 2). \tag{25}$$

For small n

$n = 2p$
$$\kappa_1 = \tfrac{1}{2}\log n-\log 2-\left(\frac{1}{2}+\frac{1}{4}+\dots+\frac{1}{n-2}\right),$$

$n = 2p+1$
$$\kappa_1 = \tfrac{1}{2}\log n-\left(1+\frac{1}{3}+\frac{1}{5}+\dots+\frac{1}{n-2}\right). \tag{26}$$

For large n

$$\kappa_1 \sim -\tfrac{1}{2}(\gamma+\log 2)+\frac{1}{2n}+\sum_{j=1}^{\infty}\frac{(-4)^{j-1}B_j}{jn^{2j}}. \tag{27}$$

Also $$2^s\kappa_s = (s-1)!\{\zeta(s,\tfrac{1}{2}n)+(-1)^s\zeta(s,\tfrac{1}{2})\} \quad (s>1), \tag{28}$$
$$= \psi^{(s-1)}(\tfrac{1}{2})+(-1)^s\psi^{(s-1)}(\tfrac{1}{2}n) \qquad (\psi^{(s-1)}(\tfrac{1}{2}) = (-1)^s(s-1)!(2^s-1)\zeta(s)). \tag{29}$$

For small n we have the following cases:

Even cumulants $(r = 2s,\ s>0)$

$$n = 2p \qquad \kappa_r = (r-1)!\left\{\zeta(r)-\left(\frac{1}{2^r}+\frac{1}{4^r}+\ldots+\frac{1}{(n-2)^r}\right)\right\}, \tag{30}$$

$$n = 2p+1 \qquad \kappa_r = (r-1)!\left\{2\zeta(r)(1-2^{-r})-\left(1+\frac{1}{3^r}+\frac{1}{5^r}+\ldots+\frac{1}{(n-2)^r}\right)\right\}.$$

Odd cumulants $(r = 2s+1,\ s>0)$

$$n = 2p \qquad \kappa_r = -(r-1)!\left\{\zeta(r)(1-2^{1-r})+\left(\frac{1}{2^r}+\frac{1}{4^r}+\ldots+\frac{1}{(n-2)^r}\right)\right\}, \tag{31}$$

$$n = 2p+1 \qquad \kappa_r = -(r-1)!\left(1+\frac{1}{3^r}+\frac{1}{5^r}+\ldots+\frac{1}{(n-2)^r}\right).$$

For large n

$$\kappa_s \sim (-1)^s(s-1)!\zeta(s)(1-2^{-s})+\frac{(s-2)!}{2n^{s-1}}+\frac{(s-1)!}{2n^s}+\frac{2}{n^s}\sum_{j=1}^{\infty}\frac{(-4)^{j-1}B_j(2j+s-2)!}{(2j)!\,n^{2j-1}} \quad (s>1). \tag{32}$$

In the special case of $n = \infty$ we have for the distribution of $\log x$, where x is a normal variable with zero mean and unit standard deviation:

$$\kappa_1 = -\tfrac{1}{2}(\gamma+\log 2), \quad \kappa_s = (-1)^s(s-1)!\zeta(s)(1-2^{-s}), \tag{33}$$

as follows also from the case of $\frac{1}{2}\log(\chi^2/n)$ on putting $n = 1$.

4. NOTE ON THE χ^2 DISTRIBUTION APPROXIMATION

Fisher's result that $\sqrt{(2\chi^2)}$ is approximately normally distributed about a mean of $\sqrt{(2n-1)}$ with unit variance (n being the number of degrees of freedom) is well known. The demonstration depends on showing that the mean value of χ is

$$\kappa_1 = \sqrt{2}\,\Gamma\tfrac{1}{2}(n+1)/\Gamma(\tfrac{1}{2}n) \sim \sqrt{(n-\tfrac{1}{2})} \quad \text{for large } n$$

and that the variance is $$n-\kappa_1^2 \sim \tfrac{1}{2},$$

but to this order of approximation it is not possible to show that γ_1 and γ_2 tend to zero with increasing n. A formula for the ratio of the two Gamma functions, developed as far as terms in n^{-3} (see Wishart (1925)), gives $\gamma_1 \sim (2n)^{-\frac{1}{2}}$ and $\gamma_2 = O(n^{-2})$ (see, for example, Kendall (1945)), but owing to the vanishing of the term in n^{-1} of γ_2 its leading term has so far not been accurately obtained, although the exact (but somewhat complicated) expressions for the β_1 and β_2 of the distribution of $s = \sigma\chi/\sqrt{(n+1)}$ were given in an editorial in *Biometrika* (1915), **10**, 522.

Since $$\tfrac{1}{2}\{\psi\tfrac{1}{2}(n+1)-\psi(\tfrac{1}{2}n)\} = \frac{1}{2n}+\sum_{j=1}^{\infty}\frac{(-1)^{j-1}(2^{2j}-1)B_j}{2jn^{2j}},$$

by the formula given in § 2, we find on integration, and insertion of the appropriate constant, that

$$\log\Gamma\tfrac{1}{2}(n+1)-\log\Gamma(\tfrac{1}{2}n) = \tfrac{1}{2}\log(\tfrac{1}{2}n)+\sum_{j=1}^{\infty}\frac{(-1)^j(2^{2j}-1)B_j}{2j(2j-1)\,n^{2j-1}},$$

and thus have $$\frac{\Gamma\frac{1}{2}(n+1)}{\Gamma(\frac{1}{2}n)} = \sqrt{(\tfrac{1}{2}n)}\exp-\frac{1}{4n}\left\{1+\sum_{j=2}^{\infty}\frac{(-1)^{j-1}(2^{2j}-1)B_j}{j(j-\frac{1}{2})\,n^{2(j-1)}}\right\}, \tag{34}$$

which can readily be expanded to give the additional terms necessary to enable the cumulants of χ (or of $\sqrt{(2\chi^2)}$) to be worked out (see Johnson & Welch (1939)). Taking $\sqrt{\{\frac{1}{2}(n-\frac{1}{2})\}}$ as the first approximation we find

$$\frac{\Gamma\frac{1}{2}(n+1)}{\Gamma(\frac{1}{2}n)} = \sqrt{\left(\frac{n-\frac{1}{2}}{2}\right)} \exp\left\{\frac{1}{16n^2}\left(1+\frac{1}{n}+\frac{1}{8n^2}-\frac{3}{4n^3}+\dots\right)\right\}$$

$$= \sqrt{\left(\frac{n-\frac{1}{2}}{2}\right)}\left(1+\frac{1}{16n(n-1)}\right)+O(n^{-3\cdot5}), \tag{35}$$

thus providing a second approximation to the ratio of two Gamma functions differing by one-half. The cumulants of $\sqrt{(2\chi^2)}$ are

$$\kappa_1 = \sqrt{(2n-1)}\left(1+\frac{1}{16n(n-1)}\right)+O(n^{-3\cdot5}),$$

$$\kappa_2 = 1-\frac{1}{4n}-\frac{1}{8n^2}+\frac{5}{64n^3}-O(n^{-4}),$$

$$\kappa_3 = \frac{1}{\sqrt{(2n)}}\left(1+\frac{1}{4n}-\frac{13}{32n^2}\right)+O(n^{-3\cdot5}), \tag{36}$$

$$\kappa_4 = \frac{3}{4n^2}\left(1+\frac{1}{n}\right)+O(n^{-4}),$$

so that

$$\gamma_1 = \frac{1}{\sqrt{(2n)}}\left(1+\frac{5}{8n}-\frac{1}{128n^2}\right)+O(n^{-3\cdot5}),$$

$$\gamma_2 = \frac{3}{4n^2}\left(1+\frac{3}{2n}\right)+O(n^{-4}).$$

The Editorial in *Biometrika* (1915), **10**, 523 calls attention in a footnote to 'Student's' approximations for the β_1 and β_2 of the sample standard deviation. The above formulae show that 'Student's' results should be

$$\beta_1 = \frac{1}{2n}\left(1+\frac{9}{4n}+\frac{31}{8n^2}\right)+O(n^{-4}),$$

$$\beta_2 = 3\left(1+\frac{1}{4n^2}+\frac{7}{8n^3}\right)+O(n^{-4}),$$

in which n is now the size of the sample. For $n = 10$ these give values too low by 2 and 5 respectively in the fourth place of decimals. Practically four-figure accuracy can be attained with n as low as 10 if in the terms in n^{-3} we replace 31/8 by 17/4 and 7/8 by 1.

REFERENCES

Bartlett, M. S. & Kendall, D. G. (1946). *J.R. Statist. Soc. Suppl.* **8**, 128.
British Association (1931). *Mathematical Tables*, **1**, 42. London: Cambridge University Press.
Cornish, E. A. & Fisher, R. A. (1937). *Rev. Inst. Internat. Statist.* **4**, 1.
Davis, H. T. (1933, 1935). *Tables of the Higher Mathematical Functions*, **1**, **2**. Indiana: Principia Press.
Fisher, R. A. & Yates, F. (1943). *Statistical Tables*, 2nd ed. London: Oliver and Boyd.
Johnson, N. L. & Welch, B. L. (1939). *Biometrika*, **31**, 216.
Kendall, M. G. (1945). *Advanced Theory of Statistics*, **1**, 2nd ed., § 12·7. London: Griffin and Co.
Pairman, E. (1919). *Tracts for Computers*, no. 1. London: Cambridge University Press.
Thompson, C. M. (1941). *Biometrika*, **32**, 168.
Wishart, J. (1925). *Biometrika*, **17**, 68.

THE MEANING OF A SIGNIFICANCE LEVEL

By G. A. BARNARD

A level of significance is a probability. To say that a given result is significant on the 5 % level means that some class of events has probability 0·05. Now whatever theory we may hold as to the nature of probability, in order to give a statement of probability a precise meaning we must refer to some reference class, or set of data, on which the probability is calculated. What is the reference class involved in a level of significance?

To many people the answer to this question seems simple enough. The reference class involved is the set of indefinite (possibly imaginary) repetitions of the experiment which gave the result in question. Otherwise put, the data, on which the probability is calculated, are the external conditions of the experiment. The following example indicates, however, that the meaning of this reference class is not always clear. The example is a modified form of one given by Prof. R. A. Fisher in a letter to the author.

Suppose we have a bag of chrysanthemum seeds, known to give plants having white flowers or plants having purple flowers, no other colours being possible. We suspect that the proportions of white and purple seeds are equal, and to test this hypothesis we select at random ten seeds from the bag, and plant them. Nine of the plants grow to maturity, and all of them have white flowers. On what level of significance can we reject the hypothesis of equality of proportions? We may assume that white and purple plants are equally viable.

It would be natural to argue that, if white and purple flowers were equally likely, the probability of our result would be $1/2^9$. If there is no reason to suspect an excess of white rather than an excess of purple flowers, we must add to this the probability of getting nine purple flowers, which is also $1/2^9$, giving a total probability of $1/2^8$. The hypothesis of equality of proportions would then be rejected on the 1/256, or the 0·3906 % level of significance. But if we did this our reference class would not be the set of indefinite repetitions of the experiment, in its ordinary meaning.

A repetition of the experiment, in its ordinary meaning, would consist of another selection of ten seeds from the bag, and their planting and growth. On such another occasion all ten plants might grow to maturity, or all or some might die. These possibilities have not been taken into account in our calculation of probability, so far.

To allow for the possible variation in the number of plants which grow, we might lay out the set of all possible results of the experiment as in Fig. 1, where n denotes the number of plants that grow, and r denotes the excess of white over purple. Thus any point in the figure can be referred to uniquely by its co-ordinates (n, r). If we now introduce a parameter p, to denote the probability (if it exists) that a plant will grow to maturity, given that it has been selected, the probability associated with the point (n, r) on the hypothesis of equality of proportions of white and purple will be

$$W(n,r;p) = \frac{10!}{n!(10-n)!}p^n(1-p)^{10-n}\frac{n!\,2^{-n}}{(\frac{1}{2}(n+r))!\,(\frac{1}{2}(n-r))!},$$

and since this is a function of the unknown p, we have a special problem of arranging the points (n, r) in order of significance before we can establish a test. The situation in this respect is similar to that dealt with in the paper on 2×2 tables, printed earlier in this issue (Barnard, 1946, pp. 123–38 above).

Proceeding as in the earlier paper, we notice first that the same level of significance must apply to (n, r) as to $(n, -r)$, so that we can confine our further considerations to the upper half of the diagram. Now in this half, the transition from (n, r) to $(n+1, r+1)$ means we discover that one of the plants which failed to grow in our case, was in fact a white-flowered plant. In this case our conviction that there is an excess of white-flowered plants would be strengthened, so that $(n+1, r+1)$ would be reckoned more significant than (n, r). Similarly, going from (n, r) to $(n+1, r-1)$ would mean that a missing plant was found to be purple, and this would weaken our belief in an excess of white-flowered plants; consequently,

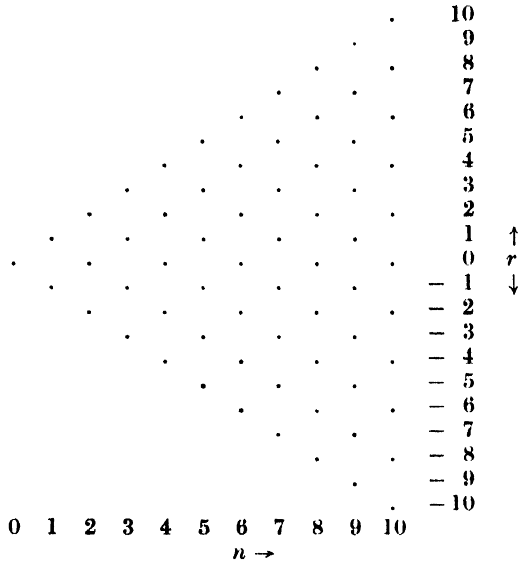


Fig. 1

(n, r) would be reckoned more significant than $(n+1, r-1)$. Finally, going from (n, r) to $(n+2, r)$ would mean growing two more plants, one purple and one white, and this would increase our tendency to believe in the equality of proportions. Consequently, (n, r) would be reckoned more significant than $(n+2, r)$. These principles taken together imply that points lying north-east, or west, of a given point (n, r), or between these two directions, would be reckoned more significant than (n, r); while, conversely, points lying east to south-west (inclusive) from (n, r) would be reckoned less significant than (n, r). The relative significance of points lying inside the half-quadrants north-east to east and south-west to west would remain undetermined.

We could now proceed as in the paper (1), building up a test, consistent with the above partial ordering, in such a way as to make the significance or otherwise of our result depend as little as possible on any knowledge we may have about the value of p. But we need not carry this through for the result we have quoted, since our conditions by themselves require that the only points in the diagram which should be reckoned not less significant than our result are the points (9, 9), (9, −9), (10, 10) and (10, −10). The probability associated with these four points is

$$P(9, 9; p) = 2(10p^9(1-p) \,.\, 2^{-9} + p^{10}2^{-10})$$
$$= (p/2)^9\,(20-19p),$$

the maximum value of which occurs when $p = 18/19$, and is $P_m(9, 9) = 0{\cdot}002413$. Thus on this basis we should conclude that our result was significant on the $0{\cdot}2413\,\%$ level.

The difference between the first result, 0·3906 %, and the second, 0·2413 %, is in practice negligible. Somewhat larger differences will be found in other similar cases, however, and it seems worth while to try to clarify the cause of the discrepancy.

Consider three possible causes for the failure of the tenth plant to grow to maturity:

(1) The bag from which the seed was taken is known to contain a proportion of dead seeds, which are physically indistinguishable from the live ones, and the tenth seed planted happened to be one of these. The conditions of growth were such that any live seed planted would have grown.

(2) The tenth plant happened to be attacked by a soil pest, which destroyed it.

(3) The statistician trod on the tenth plant while running for a bus; otherwise, it would have grown.

If we now consider what would happen in these three cases if the experiment were repeated, in case (1) we should be just as uncertain as before how many plants would grow, out of those selected. In case (2), we might or might not happen to strike a good year for the pest in question, so that we might or might not have a similar accident recurring. In case (3) we should obviously give the statistician firm instructions not to be careless, and then we could be reasonably certain that all the plants selected would grow.*

In the first case, we can suppose that the proportions of white, purple, and dead seeds in the bag are, respectively, p_1, p_2, and $1-(p_1+p_2)$; and the purpose of our experiment is to test the hypothesis $p_1 = p_2$. In this case, putting $p_1+p_2 = p$, we can clearly apply the analysis of Fig. 1, and the appropriate level of significance is 0·2413 %.

In the third case, the situation actually realized is just what it would have been if we had warned the statistician beforehand, and then thrown one of the ten seeds back into the bag. Thus our effective sample size here is 9, and the appropriate level of significance is 0·3906 %.

In the second case, the answer depends on our attitude to the set of accidents of which the pest is a specimen. If this set of accidents is regarded as a stable set of chance causes we may be justified in representing its effect on the growth of our plants by the probability p. If, on the other hand, the incidence of such pests undergoes, say, regular cyclical fluctuations from year to year, so that its incidence is to some extent predictable, if not wholly controllable, then we should not be justified in assuming the existence of a real probability corresponding to our parameter p. We should, to be on the safe side, in this case allow for the possibility that experimental technique might improve in the future, to such an extent as to eliminate the possibility of such accidents. Thus, adopting this conservative attitude to our results, we should here treat the effective sample size as 9. The repetitions of the experiment which we have in mind would then be imaginary repetitions, in which experimental technique was supposed to be better than it is now, and we have as much control over pests as we have over statisticians.

The general situation illustrated by this example can be described in terms of the notion of 'isolate' introduced by Prof. H. Levy (1931). In making an experiment, we try to construct an isolate—a system, or part of the world, which we suppose has relatively little interaction with the rest of the world, and which, for practical purposes, may be considered on its own. This isolate may contain within itself all the systems of chance causes which are

* It is not suggested that the three cases exhaust the multiplicity of types which might arise in practice. As Prof. Pearson has pointed out, if it were not the statistician, but his three-year-old son who was the vandal in case (3), we should have here a situation intermediate between our second and third instances.

regarded as affecting, to any practical extent, the results of the experiment. Such is the case in (1), where all the chance causes involved in the experiment are supposed given in the bag which is the subject of the experiment. Here, then, we are dealing with a 'good isolate', whose interaction with the rest of the world is really negligible, and chance causes operate within the isolate.

In case (3), on the other hand, we are dealing with an imperfect isolate. The outside world, in the shape of the statistician, interacts with our isolate to an extent not negligible in practice. Fortunately, in this case we are able to construct a smaller isolate, consisting of the nine surviving plants, in which the interactions with the outside world are negligible. In case (2), there may be some doubt as to what isolate we are discussing. If we regard soil pests and such things as included in the isolate, and represent them as a stable set of chance causes, then we are entitled to analyse as in case (1); but if the pests are not included in the isolate, we should analyse as in case (3).

Statistical tests are applicable to at least two types of experiment. First, to experiments in which the isolate studied contains within itself a system of chance causes which may influence the results. And second, to experiments in which the isolate studied is not a 'good' isolate, and the residual interactions with the rest of the world may affect the results. There may also be mixed cases.

The distinction between the two types may also be brought out in relation to the necessity or otherwise of an 'artificial' randomization procedure, using random digits or the like. In the first type, such an artificial randomization procedure is not strictly necessary; for example, with our bag of seeds, the bag itself, and its physically indistinguishable contents, forms a perfectly adequate randomizer. We have in this case, as it were, an impermeable shield around the system, which prevents any external shocks from affecting the system. In the second type of experiment, we need to ensure that the interactions with the outside world will not mask the results we are interested in: and if we cannot ensure a practically complete separation from the outside world, then the effect of external intereactions must be randomized, by a special procedure. The randomization here acts like a shock absorber, specially placed around the experiment to distribute external shocks evenly through the system.

In the first type of experiment, the reference class to which the significance level applies is in fact the set of indefinite repetitions of the experiment in question. In the second type of experiment, the reference class is an ideal set, in which the accidental influences of the outside world repeat themselves exactly, while the effect of these accidents on the system varies as a result of the special randomization.

REFERENCES

BARNARD, G. A. (1946). Significance tests for 2×2 tables. *Biometrika*, **34**, 123.

LEVY, H. (1931). *The Universe of Science*. London: Watts and Co.

ON THE DISTRIBUTION OF THE RANK CORRELATION COEFFICIENT τ WHEN THE VARIATES ARE NOT INDEPENDENT

BY WASSILY HÖFFDING

I. INTRODUCTION

1. Consider a population distributed according to two variates x, y. Two members (x_1, y_1) and (x_2, y_2) of the population will be called *concordant* if both values of one member are greater than the corresponding values of the other one, that is if

$$x_1 < x_2,\ y_1 < y_2 \quad \text{or} \quad x_1 > x_2,\ y_1 > y_2.$$

They will be called *discordant* if for one member one value is greater and the other one smaller than for the other member, that is if

$$x_1 < x_2,\ y_1 > y_2 \quad \text{or} \quad x_1 > x_2,\ y_1 < y_2.$$

The probability p that two members drawn from the population at random without replacement are concordant will be called the *probability of concordance*, the probability q that they are discordant will be called the *probability of discordance*.

In the following only populations will be considered for which the probabilities of $x_1 = x_2$ or $y_1 = y_2$ are zero, so that

$$p+q = 1. \tag{1}$$

The main types of such populations are (*a*) an infinite population with both x and y distributed continuously, (*b*) a finite population where all values of x and all values of y are different among themselves. The condition that the two members are drawn without replacement is, of course, only relevant in case (*b*).

For a sample of n members drawn from the population, the probabilities of concordance and discordance are defined in the same manner as for the population. They will be denoted by p' and q' to distinguish them from the population values. If for the population (1) is fulfilled, it may be assumed that all values of x and all values of y in the sample are different, so that

$$p'+q' = 1. \tag{2}$$

It follows from the definition that p' is the relative frequency of concordant pairs among the $\binom{n}{2}$ pairs which can be formed from the members of the sample.

The probability of concordance expresses an essential property of a bivariate distribution. It may in itself be considered as a measure of correlation. p' is an estimate of p; it will be shown that the mean value of p' is p. If a coefficient lying between the limits -1 and $+1$ is preferred, the quantity

$$\tau = p'-q' = 2p'-1 \tag{3}$$

may be taken.

2. The quantity p here termed the probability of concordance was apparently first considered by Esscher (1924) who also used the quantity

$$D = \frac{1}{\binom{n}{2}} \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} \operatorname{sign}(x_i - x_j) \operatorname{sign}(y_i - y_j),$$

(where x_i, y_i, $i = 1, \ldots, n$, are the sample values of the variates) which is the same as the coefficient τ as defined by (3). Esscher showed that if x and y are normally correlated with correlation coefficient r, the expectation of $D = \tau$ is

$$E(\tau) = \frac{2}{\pi} \sin^{-1} r. \tag{4}$$

Hence, from this equation, he suggested estimating r from ranked data by means of the relation

$$r = \sin \frac{\pi}{2} \tau = \sin \pi \left(p' - \tfrac{1}{2}\right).$$

For the variance of τ Esscher found in the case of a normally distributed population

$$\frac{1}{4}\binom{n}{2} \operatorname{var}(\tau) = pq + \frac{n-2}{2}\left\{\frac{1}{9} - \left(\frac{2}{\pi} \sin^{-1}\left(\frac{r}{2}\right)\right)^2\right\}, \tag{5}$$

where
$$4pq = 1 - \left(\frac{2}{\pi} \sin^{-1} r\right)^2.$$

While Esscher saw in p' and $D = \tau$ only a means for estimating r, Lindeberg (1926) stressed the significance of the probability of concordance itself for judging the degree of dependence between the variates. He proposed for that purpose the coefficient

$$P = 100p' - 50 = 50\tau,$$

called by him *Korrelationsprozent*. Lindeberg also gave, without proof, a formula for the variance of p' in the general case of correlated variates (see (13) below).

Jordan (1927) suggested using, instead of Lindeberg's P, the coefficient later termed by Kendall τ.

Kendall (1938), independently of the above authors, proposed τ as a measure of rank correlation. He completely solved the problem of the sampling distribution of τ in a universe in which all possible rankings are equally probable, showing that it rapidly tends to normality for increasing n.

3. The main object of this paper is to show that the sampling distribution of p' (and hence that of τ) tends to normality as $n \to \infty$ for any population with continuously distributed x and y if a certain condition is fulfilled (Part IV). In addition, Lindeberg's formula for the variance of p' is proved (Part II) and extended for a finite population (Part V). Finally, in Part VI the problem of estimating var (p') from the sample is considered.

II. Mean value and variance of p' in the case of an infinite population

4. Consider a sample of n drawn at random from an infinite population with continuous x and y. Replace the values of x and of y in the sample by their ranks and arrange the members of the sample so that the ranking of x is $1, 2, \ldots, n$. Then the ranking of y is a permutation

$$\Pi = (\pi_1, \ldots, \pi_n)$$

of the numbers $1, \ldots, n$.

Let I and J be the numbers of inversions in the permutations $(\pi_n, \pi_{n-1}, \ldots, \pi_1)$ and $(\pi_1, \ldots, \pi_n)$. Then

$$p' = \frac{2I}{n(n-1)}, \quad q' = \frac{2J}{n(n-1)}. \tag{6}$$

Thus the knowledge of the permutation Π corresponding to the given sample is sufficient for evaluating p'.

5. Let $P(\Pi)$ be the probability of drawing a random sample represented by the permutation Π. Let $p'(\Pi)$ be the probability of concordance for such a sample. Then

$$p = \Sigma P(\Pi) p'(\Pi), \tag{7}$$

where the sum is extended over all permutations Π of n numbers.

The right-hand side of (7) is equal to the mean value of p'. Hence

$$Ep' = p. \tag{8}$$

Consider, in generalization of p', the probability w' that among $m \leqslant n$ members drawn from the sample at random without replacement, certain pairs of members are concordant; for instance, among four members A, B, C, D, the pairs AB, AC, AD; or the pairs AB, CD, etc. Let w be the corresponding probability for the parent population. Then it is seen in the same manner as with p' that

$$Ew' = w. \tag{9}$$

Thus, if we can express $(p')^\mu$, the probability of drawing μ concordant pairs from the sample, replacing each pair after drawing it, by probabilities without replacement of the type w', we can also, in virtue of (9), represent $E(p')^\mu$ by population parameters of the type w.

6. Now, $(p')^2$, the probability of drawing from the sample one concordant pair and, after replacing it, of drawing again a concordant pair, is the sum of the following three probabilities:

(*a*) the probability of getting the same pair in both drawings $\left(1\Big/\binom{n}{2}\right)$, multiplied by the probability that this pair is concordant (p');

(*b*) the probability that the second pair has one member in common with the first pair $\left(2(n-2)\Big/\binom{n}{2}\right)$, multiplied by the probability, say k', that among three members A, B, C drawn from the sample without replacement, one, say A, is concordant with the other two;

(*c*) the probability that the second pair has no member in common with the first one $\left(\binom{n-2}{2}\Big/\binom{n}{2}\right)$, multiplied by the probability that among four members A, B, C, D drawn without replacement, two pairs without a member in common, say AB and CD, are concordant. The latter probability may be denoted by $(p^2)'$ since the corresponding probability for the infinite population is p^2.

Thus,
$$\binom{n}{2}(p')^2 = p' + 2(n-2)k' + \binom{n-2}{2}(p^2)', \tag{10}$$

and, applying (9),
$$\binom{n}{2}E(p')^2 = p + 2(n-2)k + \binom{n-2}{2}p^2. \tag{11}$$

Hence, we have for the variance of p'

$$\binom{n}{2}\operatorname{var}(p') = \binom{n}{2}\{E(p')^2 - p^2\} = p + 2(n-2)k - (2n-3)p^2 \tag{12}$$

or
$$\binom{n}{2}\operatorname{var}(p') = \binom{n}{2}\mu_2(p') = p(1-p) + 2(n-2)(k-p^2). \tag{13}$$

This is identical with the formula given without proof by Lindeberg (1926).

7. In the case considered by Kendall where all permutations Π of n numbers are equally probable, the permutations of $m \leqslant n$ also are equiprobable. Hence

$$p = P(1,2) = q = P(2,1) = \tfrac{1}{2}.$$

Further, representing k as the mean value of k' in a sample of 3, we find

$$k = P(123)+\tfrac{1}{3}P(132)+\tfrac{1}{3}P(213) = (1+\tfrac{1}{3}+\tfrac{1}{3})\frac{1}{3!} = \frac{5}{18}.$$

Inserting these values in (13), we have

$$\operatorname{var}(p') = \frac{2n+5}{18n(n-1)},$$

$$\operatorname{var}(\tau) = 4\operatorname{var}(p') = \frac{2(2n+5)}{9n(n-1)}$$

in accordance with Kendall's formula.

III. Some algebraic formulae

8. We shall now consider some algebraic relations to be used in the proof of normality of p' for large n.

Let $f_d(\rho)$ be a polynomial of degree d in ρ. Then

$$\sum_{\rho=0}^{\beta}(-1)^{\beta-\rho}\binom{\beta}{\rho}f_d(\rho) = \begin{cases} 0 & \text{if} \quad d<\beta, \\ a_0\beta! & \text{if} \quad d=\beta, \end{cases} \tag{14}$$

where a_0 is the coefficient of the highest power ρ^d in $f_d(\rho)$.

To prove (14) write $\qquad f_d(\rho) = a_0\rho^{[d]}+a_1\rho^{[d-1]}+\dots,$

where $\qquad \rho^{[0]} = 1, \quad \rho^{[\delta]} = \rho(\rho-1)\dots(\rho-\delta+1), \quad (\delta \geqslant 1).$

Then (14) follows from the fact that

$$\sum_{\rho=0}^{\beta}(-1)^{\beta-\rho}\binom{\beta}{\rho}\rho^{[\delta]} = \beta^{[\delta]}\sum_{\rho-\delta=0}^{\beta-\delta}(-1)^{\beta-\rho}\binom{\beta-\delta}{\rho-\delta}$$

is equal to $\beta^{[\delta]}(1-1)^{\beta-\delta} = 0$ if $\beta-\delta>0$ and to $\beta!$ if $\beta=\delta$.

9. For any non-negative integer ν we may write

$$n^\nu = d^{(\alpha)}_{\nu 0}(n-\alpha)^{[\nu]}+d^{(\alpha)}_{\nu 1}(n-\alpha)^{[\nu-1]}+\dots+d^{(\alpha)}_{\nu,\nu-1}(n-\alpha)+d^{(\alpha)}_{\nu\nu}. \tag{15}$$

We will study certain properties of the coefficients $d^{(\alpha)}_{\nu\kappa}$.

From (15) it is seen immediately that

$$d^{(\alpha)}_{\nu 0} = 1. \tag{16}$$

Inserting in (15) $n=\alpha+\beta$ ($\beta = 0, 1, \dots$) we have

$$(\alpha+\beta)^\nu = d^{(\alpha)}_{\nu,\nu-\beta}\beta!+d^{(\alpha)}_{\nu,\nu-\beta+1}\beta^{[\beta-1]}+\dots+d^{(\alpha)}_{\nu,\nu-1}\beta+d^{(\alpha)}_{\nu\nu} \tag{17}$$

or

$$d^{(\alpha)}_{\nu\nu} = \alpha^\nu, \quad d^{(\alpha)}_{\nu,\nu-\beta} = \frac{1}{\beta!}\left\{(\alpha+\beta)^\nu-\sum_{\lambda=0}^{\beta-1}\beta^{[\lambda]}d^{(\alpha)}_{\nu,\nu-\lambda}\right\}, \quad (\beta = 1, 2, \dots). \tag{18}$$

Hence we find by induction

$$d^{(\alpha)}_{\nu,\nu-\beta} = \frac{1}{\beta!}\sum_{\rho=0}^{\beta}(-1)^{\beta-\rho}\binom{\beta}{\rho}(\alpha+\rho)^\nu. \tag{19}$$

If we take this as definition of $d^{(\alpha)}_{\nu\kappa}$ for $\kappa<0$, we have in virtue of (14)

$$d^{(\alpha)}_{\nu\kappa} = 0 \quad \text{for} \quad \kappa<0. \tag{20}$$

Expanding $(\alpha+\rho)^\nu$ we have from (19)

$$d^{(\alpha)}_{\nu,\nu-\beta} = \frac{1}{\beta!}\sum_{\rho=0}^{\beta}(-1)^{\beta-\rho}\binom{\beta}{\rho}\sum_{\sigma=0}^{\nu}\binom{\nu}{\sigma}\alpha^\sigma\rho^{\nu-\sigma} = \sum_{\sigma=0}^{\nu}\binom{\nu}{\sigma}\alpha^\sigma\frac{1}{\beta!}\sum_{\rho=0}^{\beta}(-1)^{\beta-\rho}\binom{\beta}{\rho}\rho^{\nu-\sigma}.$$

Comparing the last sum with (19) and writing

$$d_{\nu\kappa} = d_{\nu\kappa}^{(0)},$$

we have

$$d_{\nu,\nu-\beta}^{(\alpha)} = \sum_{\sigma=0}^{\nu} \binom{\nu}{\sigma} d_{\nu-\sigma,\nu-\beta-\sigma}\alpha^{\sigma}$$

or, putting $\nu-\beta = \kappa$ and noting that, by (20), $d_{\nu-\sigma,\kappa-\sigma} = 0$ for $\sigma > \kappa$,

$$d_{\nu\kappa}^{(\alpha)} = \sum_{\sigma=0}^{\kappa} \binom{\nu}{\sigma} d_{\nu-\sigma,\kappa-\sigma}\alpha^{\sigma}. \tag{21}$$

We have the recurrence relation

$$d_{\nu+1,\kappa}^{(\alpha)} - d_{\nu\kappa}^{(\alpha)} = (\alpha+\nu+1-\kappa)\, d_{\nu,\kappa-1}^{(\alpha)} \tag{22}$$

which can be obtained by multiplying (15) by n, then writing down (15) with $\nu+1$ instead of ν, and comparing coefficients in both expressions.

10. We prove now two properties of the coefficients $d_{\nu\kappa}^{(\alpha)}$.

(I) $d_{\nu\kappa}$ is a polynomial in ν of degree 2κ, the term of highest degree being $\nu^{2\kappa}/2^{\kappa}\kappa!$.

In virtue of (16) this is true for $\kappa = 0$. And if it is true for $\kappa-1$, the highest term of $d_{\nu+1,\kappa} - d_{\nu\kappa}$ is, by (22) with $\alpha = 0$, $\nu^{2\kappa-1}/2^{\kappa-1}(\kappa-1)!$, and hence that of $d_{\nu\kappa}$, by a well-known theorem,

$$\frac{1}{2\kappa}\,\frac{1}{2^{\kappa-1}(\kappa-1)!}\,\nu^{2\kappa} = \frac{1}{2^{\kappa}\kappa!}\,\nu^{2\kappa}.$$

(II) $d_{t-\rho,\kappa}^{(\gamma-t)}$ is a polynomial in t of degree 2κ with the highest term $(-1)^{\kappa}t^{2\kappa}/2^{\kappa}\kappa!$.

From (21),

$$d_{t-\rho,\kappa}^{(\gamma-t)} = \sum_{\sigma=0}^{\kappa} \binom{t-\rho}{\sigma} d_{t-\rho-\sigma,\kappa-\sigma}(\gamma-t)^{\sigma}.$$

In $\binom{t-\rho}{\sigma}$ the highest term in t is $\frac{1}{\sigma!}t^{\sigma}$.

In $d_{t-\rho-\sigma,\kappa-\sigma}$ the highest term in t is $\frac{1}{2^{\kappa-\sigma}(\kappa-\sigma)!}t^{2\kappa-2\sigma}$ (by (I)).

In $(\gamma-t)^{\sigma}$ the highest term in t is $(-1)^{\sigma}t^{\sigma}$.

Hence, in $d_{t-\rho,\kappa}^{(\gamma-t)}$ the highest term is

$$\sum_{\sigma=0}^{\kappa} \frac{(-2)^{\sigma}}{2^{\kappa}\sigma!(\kappa-\sigma)!}\,t^{2\kappa} = \frac{1}{2^{\kappa}\kappa!}(1-2)^{\kappa}t^{2\kappa} = \frac{(-1)^{\kappa}}{2^{\kappa}\kappa!}\,t^{2\kappa}.$$

11. $d_{\nu,\nu-\beta}$ has also a combinatorial meaning.

Let

$$\Sigma_{\beta}(\nu) = \Sigma\,\frac{\nu!}{\nu_1!\,\nu_2!\dots\nu_{\beta}!},\quad \Sigma'_{\beta}(\nu) = \Sigma'\,\frac{\nu!}{\nu_1!\,\nu_2!\dots\nu_{\beta}!}$$

where Σ indicates summation over all $\nu_i \geqslant 0$, Σ' over all $\nu_i \geqslant 1$, and in both cases $\nu_1+\dots+\nu_{\beta} = \nu$. $\Sigma_{\beta}(\nu)$ is the number of ways of allocating ν objects on β places, and $\Sigma'_{\beta}(\nu)$ is the number of ways of allocating ν objects on β places in such a manner that no place remains empty.

We have

$$\Sigma_{\beta}(\nu) = \beta^{\nu},$$

and a little consideration shows that

$$\Sigma_{\beta}(\nu) = \Sigma'_{\beta}(\nu) + \binom{\beta}{1}\Sigma'_{\beta-1}(\nu) + \dots + \binom{\beta}{\beta-1}\Sigma'_{1}(\nu).$$

Comparing this with (17) we see that

$$d_{\nu,\nu-\beta} = \frac{1}{\beta!}\Sigma'_{\beta}(\nu) = \frac{1}{\beta!}\Sigma'\,\frac{\nu!}{\nu_1!\,\nu_2!\dots\nu_{\beta}!}. \tag{23}$$

IV. Proof of normality of p' for $n \to \infty$

12. Any set of different pairs of elements belonging to the population will be briefly referred to as a *system* (two pairs being different if they have no more than one element in common).

If we represent the elements of a system by points in a plane and the pairs of elements by lines joining the points, we have a *pattern* corresponding to the given system. Two systems will be said to have the same pattern if there exists a one-to-one correspondence between the elements of both systems such that if two elements of one system form a pair, the two corresponding elements of the other system also form a pair. Thus the only thing relevant in a pattern is the lines connecting the points, the position of the points having no significance.

A pattern will be called *simple* if one can pass from any point of the pattern to any other one along lines belonging to the pattern. A *composite* pattern is a pattern consisting of more than one simple pattern.

If the elements of a system (or the points of the corresponding pattern) are denoted by different letters A, B, C, ..., each pair of the system can be represented by a pair of letters. All systems of one pair have the same pattern (AB). There are two patterns of two pairs, one simple and containing three points (AB, BC) and one composite and containing four points (AB, CD). There are five patterns of three pairs, three simple $(AB, BC, CA; AB, BC, CD; AB, AC, AD)$, one consisting of two different simple patterns (AB, CD, DE) and one consisting of three equal simple patterns (AB, CD, EF).

13. If a *simple* pattern consists of a_j points and b_j pairs,

$$a_j \leqslant b_j + 1. \tag{24}$$

For this is true for $b_j = 1$, and by adding one pair to a simple pattern, at most one point is added if the new pattern is to be simple again.

Denote the different simple patterns by $S_1, S_2, \ldots, S_j, \ldots$, where S_1 stands for the one-pair pattern and S_2 for the two-pairs pattern (AB, BC), all S_j with $j \geqslant 3$ consisting of three or more pairs. Let a_j be the number of points and b_j the number of pairs in S_j. Then $a_1 = 2$, $a_2 = 3$,

$$b_1 = 1, \quad b_2 = 2, \quad b_j \geqslant 3 \quad \text{if} \quad j \geqslant 3. \tag{25}$$

Consider a pattern P composed of γ_1 simple patterns S_1, γ_2 simple patterns S_2, etc., and containing a points and b pairs. Then, writing symbolically

$$P = \Sigma \gamma_j S_j,$$

we have

$$a = \Sigma \gamma_j a_j, \quad b = \Sigma \gamma_j b_j.$$

In virtue of (24),

$$3b - 2a = \Sigma \gamma_j (3b_j - 2a_j) \geqslant \Sigma \gamma_j (b_j - 2),$$

and from (25)

$$3b - 2a \geqslant -\gamma_1, \tag{26}$$

the sign of equality holding if, and only if, pattern P contains no other simple patterns than S_1 and S_2.

14. $(p')^\mu$ is the probability that μ pairs of elements drawn from the sample, replacing each pair after drawing, are all concordant. We may write

$$(p')^\mu = \Sigma A_i w_i',$$

where A_i is the probability that μ pairs are drawn from the sample in such a way that the system of different pairs among them has the pattern P_i, and if a_i is the number of points in P_i, w_i' is the probability that if a_i elements are drawn from the sample without replacement

and paired according to pattern P_i, all pairs of P_i are concordant. The summation is extended over all patterns P_i with no more than μ pairs.

Since the probabilities w_i' are of the type for which formula (9) is applicable, we have

$$E(p')^{\mu} = \Sigma A_i w_i, \tag{27}$$

where, as usual, w_i is the population probability corresponding to the sample probability w_i'.

15. Consider a term Aw in (27) corresponding to the pattern

$$P = \Sigma \gamma_j S_j$$

with $\gamma = \Sigma \gamma_j$ simple patterns, $a = \Sigma \gamma_j a_j$ points and $b = \Sigma \gamma_j b_j$ pairs.

Let
$$\bar{P} = \sum_{j \geq 2} \gamma_j S_j$$

be the pattern obtained from P by excluding the single-pair patterns S_1. Then

$$\bar{\gamma} = \gamma - \gamma_1, \quad \bar{a} = a - 2\gamma_1 \quad \text{and} \quad \bar{b} = b - \gamma_1 \tag{28}$$

are the numbers of simple patterns, points and pairs in $\bar{P}$.

We have
$$w = p^{\gamma_1} v, \tag{29}$$

where v is independent of p and γ_1, only depending on the pattern $\bar{P}$.

16. The probability A will be studied, in the first place, as a function of n and γ_1, while its dependence on $\bar{P}$ will be considered later and only in a special case. It must be borne in mind that, by (28), γ, a and b also depend on γ_1.

Let $Q_1, Q_2, \ldots, Q_b$ be the pairs of pattern P numbered in some definite order. Suppose pair Q_β appears μ_β times ($\beta = 1, \ldots, b$). Then

$$\mu_1 + \ldots + \mu_b = \mu, \quad \mu_\beta \geq 1 \quad (\beta = 1, \ldots, b).$$

Let $R_1, R_2, \ldots, R_\mu$ be the total set of the pairs drawn, numbered independently of the order in which they appear. Then μ_1 R's are equal to Q_1, μ_2 R's are equal to Q_2, etc.

Let B be the probability that among μ pairs drawn from the sample, replacing each pair after drawing, b pairs are different and arranged according to pattern P, pair Q_β appearing μ_β times ($\beta = 1, \ldots, b$) and the μ pairs being drawn in a definite order, say $R_1, R_2, \ldots, R_\mu$.

Suppose, R_1 is a Q_1. Since any pair drawn may be taken as Q_1 (only the relative position of the pairs being relevant), the probability first to draw R_1 is 1. The probability that the second pair drawn is R_2 depends on whether R_2 has no, one or both elements in common with R_1. In the first case, it is $\binom{n-2}{2}\Big/\binom{n}{2}$, in the second case $2(n-2)\Big/\binom{n}{2}$ (the factor 2 arising from the fact that each of the two elements of R_1 can be the element common with R_2), and in the third case, $1\Big/\binom{n}{2}$.

In general, if the first λ pairs drawn are $R_1, \ldots, R_\lambda$, and if they form a pattern P' containing α different elements, the probability that the $(\lambda+1)$th pair drawn is $R_{\lambda+1}$ depends on whether $R_{\lambda+1}$ has no, one or both elements in common with P'. In the first case it is $\binom{n-\alpha}{2}\Big/\binom{n}{2}$, in the second case, $c'(n-\alpha)\Big/\binom{n}{2}$, and in the third case, $c''\Big/\binom{n}{2}$, where c' and c'' are independent of n. If, in the last case, $R_{\lambda+1}$ is equal to one of the preceding R's, $c'' = 1$.

B is the product of all μ such probabilities, and it is seen from the above consideration that it is of the form

$$B = C(n-2)^{[a-2]}\binom{n}{2}^{-\mu+1},$$

where C is independent of n.

We also see that a pair which has already appeared before makes no contribution to C. Hence, C only depends on the different pairs of pattern P, and is independent of the numbers μ_β.

The above reflexion further shows that for any simple pattern contained in P, the pair drawn first, having no elements in common with the preceding pairs, contributes to C the factor $\frac{1}{2}$, except for the first pair, R_1, which yields the factor 1. Thus, C contains the factor $2^{-\gamma+1} = 2^{-\gamma_1-\bar{\gamma}+1}$, and $2^{-\gamma_1}$ is obviously the sole contribution to C from the γ_1 single pairs (pattern S_1) contained in P. Hence

$$B = 2^{-\gamma_1}C'(n-2)^{[a-2]}\binom{n}{2}^{-\mu+1},$$

where C' is independent of n and γ_1, and also independent of the order in which the γ_1 single pairs are drawn.

A, the probability that μ pairs drawn form pattern P, irrespective of the order in which they appear, depends on n in the same way as B. As a function of γ_1, A contains, besides $2^{-\gamma_1}$, the factor $1/\gamma_1!$ owing to the fact that the γ_1 single pairs are interchangeable. Further it contains the factor $\Sigma_b'(\mu)$ which indicates the number of ways of allocating μ objects on b places so that no place remains empty. In virtue of (23) we have

$$\Sigma_b'(\mu) = b!\,d_{\mu,\mu-b}.$$

Thus, A is of the form
$$A = D_{\gamma_1}^{(\mu)}(n-2)^{[2\gamma_1+\bar{a}-2]}\binom{n}{2}^{-\mu+1}, \tag{30}$$

where
$$D_{\gamma_1}^{(\mu)} = 2^{-\gamma_1}\frac{(\gamma_1+\bar{b})!}{\gamma_1!}d_{\mu,\mu-\gamma_1-\bar{b}}D' \tag{31}$$

and D' is independent of both n and γ_1 and only depends on the pattern $\bar{P}$ containing no S_1.

Inserting (29) and (30) in (27), we have

$$\binom{n}{2}^{\mu-1}E(p')^\mu = \Sigma\, D_{\gamma_1}^{(\mu)}p^{\gamma_1}v(n-2)^{[2\gamma_1+\bar{a}-2]}, \tag{32}$$

the summation taking place over all patterns with no more than μ pairs. (32) also holds for $\mu = 0$ if by a 'pattern of 0 pairs' we understand the case $\gamma_1 = \gamma_2 = \ldots = 0$ and take (31) as definition of $D_{\gamma_1}^{(0)}$ with suitably chosen D'. $\rho^{[-\delta]}$ with $\delta > 0$ is defined by

$$\rho^{[-\delta]}(\rho+\delta)^{[\delta]} = 1.$$

17. If
$$\mu_\nu(p') = E(p'-p)^\nu,$$

we have
$$\binom{n}{2}^{\nu-1}\mu_\nu(p') = \sum_{\delta=0}^{\nu}(-1)^\delta\binom{\nu}{\delta}\binom{n}{2}^\delta p^\delta\binom{n}{2}^{\nu-\delta-1}E(p')^{\nu-\delta}.$$

Applying (32) with
$$\mu = \nu-\delta, \quad \gamma_1 = \kappa-\delta, \tag{33}$$

we have for the coefficient of $p^\kappa v$ in $\binom{n}{2}^{\nu-1}\mu_\nu(p')$

$$\sum_{\delta=0}^{\nu}(-1)^\delta\binom{\nu}{\delta}\binom{n}{2}^\delta D_{\kappa-\delta}^{(\nu-\delta)}(n-2)^{[\bar{a}+2\kappa-2\delta-2]}$$

$$= \sum_{\delta=0}^{\nu}(-1)^\delta\binom{\nu}{\delta}\frac{1}{2^\delta}\sum_{\rho=0}^{\delta}(-1)^\rho\binom{\delta}{\rho}n^{2\delta-\rho}D_{\kappa-\delta}^{(\nu-\delta)}(n-2)^{[\bar{a}+2\kappa-2\delta-2]}.$$

Inserting here, in accordance with (15),

$$n^{2\delta-\rho} = \sum_{\sigma=0}^{2\delta-\rho} d_{2\delta-\rho,\sigma}^{(\bar{a}+2\kappa-2\delta)} (n-\bar{a}-2\kappa+2\delta)^{[2\delta-\rho-\sigma]},$$

we have $\displaystyle\sum_{\delta=0}^{\nu} (-1)^\delta \binom{\nu}{\delta} \frac{1}{2^\delta} \sum_{\rho=0}^{\delta} (-1)^\rho \binom{\delta}{\rho} D_{\kappa-\delta}^{(\nu-\delta)} \sum_{\sigma=0}^{2\delta-\rho} d_{2\delta-\rho,\sigma}^{(\bar{a}+2\kappa-2\delta)} (n-2)^{[\bar{a}+2\kappa-2-\rho-\sigma]}.$

Putting $\alpha = \bar{a}+2\kappa-2-\rho-\sigma$, we have for the coefficient $K_\kappa^{(\nu,\alpha)}$ of $p^\kappa v(n-2)^{[\alpha]}$ in $\binom{n}{2}^{\nu-1} \mu_\nu(p')$

$$K_\kappa^{(\nu,\alpha)} = \sum_{\delta=0}^{\nu} (-1)^\delta \binom{\nu}{\delta} a_\kappa^{(\nu,\alpha)}(\delta), \tag{34}$$

where
$$a_\kappa^{(\nu,\alpha)}(\delta) = \frac{1}{2^\delta} D_{\kappa-\delta}^{(\nu-\delta)} \sum_{\rho=0}^{\delta} (-1)^\rho \binom{\delta}{\rho} d_{2\delta-\rho,\bar{a}+2\kappa-\alpha-\rho-2}^{(\bar{a}+2\kappa-2\delta)}. \tag{35}$$

Since $d_{2\delta-\rho,\bar{a}+2\kappa-\alpha-\rho-2}^{(\bar{a}+2\kappa-2\delta)} = 0$ if $\bar{a}+2\kappa-\alpha-\rho-2<0$, the upper limit of ρ in the summation may be taken as $\bar{a}+2\kappa-\alpha-2$, which is independent of δ. We have then, in virtue of (31),

$$a_\kappa^{(\nu,\alpha)}(\delta) = \frac{1}{2^\kappa} (\bar{b}+\kappa-\delta)^{[\bar{b}]} d_{\nu-\delta,\nu-\bar{b}-\kappa} D' \sum_{\rho=0}^{\bar{a}+2\kappa-\alpha-2} (-1)^\rho \binom{\delta}{\rho} d_{2\delta-\rho,\bar{a}+2\kappa-\alpha-\rho-2}^{(\bar{a}+2\kappa-2\delta)}, \tag{36}$$

where D' is independent of δ.

In virtue of (I) and (II), para. 10, $a_\kappa^{(\nu,\alpha)}(\delta)$ is a polynomial in δ. The degree of the $(\rho+1)$th term in the sum in (36) is $\rho+2(\bar{a}+2\kappa-\alpha-\rho-2)$, which is highest for $\rho=0$. Hence, the degree d of $a_\kappa^{(\nu,\alpha)}(\delta)$ is

$$d = \bar{b}+2(\nu-\bar{b}-\kappa)+2(\bar{a}+2\kappa-\alpha-2) = 2(\nu+\bar{a}+\kappa-\alpha-2)-\bar{b}. \tag{37}$$

Now, according to (34) and (14), $K_\kappa^{(\nu,\alpha)} = 0$ if $d<\nu$, or, in virtue of (37),

$$K_\kappa^{(\nu,\alpha)} = 0 \quad \text{if} \quad 2\alpha > \nu-4+2\bar{a}-\bar{b}+2\kappa. \tag{38}$$

Applying (26) for pattern $\bar{P}$, we have, since $\gamma_1 = 0$, $2\bar{a} \leqslant 3\bar{b}$, and consequently

$$2\bar{a}-\bar{b}+2\kappa \leqslant 2(\bar{b}+\kappa),$$

the sign of equality holding if and only if pattern P contains no other simple patterns than S_1 and S_2.

Remembering that, according to para. 16, $b \leqslant \mu$, we have in virtue of (33)

$$\bar{b}+\kappa = \bar{b}+\gamma_1+\delta = b+\delta \leqslant \mu+\delta = \nu. \tag{39}$$

Thus in any case
$$2\bar{a}-\bar{b}+2\kappa \leqslant 2\nu$$

and, in virtue of (38),
$$K_\kappa^{(\nu,\alpha)} = 0 \quad \text{if} \quad \alpha \geqslant \tfrac{3}{2}\nu-1. \tag{40}$$

If P contains at least one simple pattern with more than two pairs, we even have

$$2\bar{a}-\bar{b}+2\kappa < 2\nu,$$

and consequently
$$K_\kappa^{(\nu,\alpha)} = 0 \quad \text{if} \quad \alpha \geqslant \tfrac{3}{2}\nu-2. \tag{41}$$

From (40) it appears that the degree in n of $\binom{n}{2}^{\nu-1} \mu_\nu(p')$ is

$$\leqslant 3h-2 = \tfrac{3}{2}\nu-2 \quad \text{if} \quad \nu = 2h,$$
$$\leqslant 3h-1 = \tfrac{3}{2}\nu-\tfrac{5}{2} \quad \text{if} \quad \nu = 2h+1.$$

Thus, in $\mu_{2h+1}(p')$, if expanded in powers of n, the degree of the highest term is

$$\leqslant 3h-1-4h = -h-1.$$

In $\mu_2(p')$, in virtue of (13), the degree of the highest term is -1, provided that $k-p^2 \neq 0$. Hence, the degree of

$$\alpha_{2h+1}(p') = \frac{\mu_{2h+1}(p')}{\mu_2^{\frac{1}{2}(2h+1)}(p')}$$

is $\leqslant -h-1+h+\frac{1}{2} = -\frac{1}{2}$. It follows that

$$\alpha_{2h+1}(p') \underset{n}{\rightarrow} 0 \quad \text{if} \quad k-p^2 > 0. \tag{42}$$

($k-p^2 < 0$ is impossible since in this case var (p') would become < 0 for large n.)

18. As we have seen, we may write

$$\binom{n}{2}^{2h-1} \mu_{2h}(p') = R_h(n-2)^{[3h-2]} + R'_h(n-2)^{[3h-3]} + \ldots.$$

Then it follows from (41) that R_h only contains terms depending on patterns S_1 and S_2, that is, R_h is of the form

$$R_h = \sum_{\kappa}\sum_{\lambda} K_{\kappa,\lambda}^{(2h,3h-2)} p^{\kappa} k^{\lambda}. \tag{43}$$

The only terms in $\binom{n}{2}^{\mu-1} E(p')^{\mu}$ which can contribute to this sum are of the form

$$D_{\gamma_1}^{(\mu)} p^{\gamma_1} k^{\lambda} (n-2)^{[2\gamma_1+3\lambda-2]}.$$

The pattern corresponding to such a term is

$$P = \gamma_1 S_1 + \lambda S_2.$$

Remembering the considerations in para. 16, we see that in each S_2 the pair drawn first contributes to C the factor $\frac{1}{2}$ (except if it is R_1), while the first drawing of the other pair yields the factor 2. Hence, the S_2's make no contribution to C, and we have

$$C = 2^{-\gamma_1+1}.$$

The contribution of the patterns S_2 to A is twofold: since in each S_2 the two pairs may be interchanged, this gives the factor $(1/2!)^{\lambda}$; and since the λ patterns S_2 may be interchanged, we have the factor $1/\lambda!$. Thus

$$D_{\gamma_1}^{(\mu)} = 2^{-\gamma_1-\lambda+1} \frac{(\gamma_1+2\lambda)!}{\gamma_1!\,\lambda!} d_{\mu,\mu-\gamma_1-2\lambda}$$

and, in virtue of (31), since $\bar{b} = 2\lambda$,

$$D' = \frac{1}{2^{\lambda-1}\lambda!}. \tag{44}$$

Inserting in (38) $\qquad \nu = 2h, \quad \alpha = 3h-2, \quad \bar{a} = 3\lambda, \quad \bar{b} = 2\lambda, \qquad$ (45)

we see that $K_{\kappa,\lambda}^{(2h,3h-2)} \neq 0$ is possible only if

$$\kappa + 2\lambda \geqslant 2h.$$

On the other hand, from (39), $\qquad \kappa + 2\lambda \leqslant 2h.$

Hence, $\qquad \kappa = 2h - 2\lambda. \qquad$ (46)

Inserting this in (43), we have

$$R_h = \sum_{\lambda=0}^{h} K_{2(h-\lambda),\lambda}^{(2h,3h-2)} p^{2(h-\lambda)} k^{\lambda}, \tag{47}$$

where

$$K_{2(h-\lambda),\lambda}^{(2h,3h-2)} = \sum_{\delta=0}^{2h} (-1)^{\delta} \binom{2h}{\delta} a_{2(h-\lambda)}^{(2h,3h-2)}(\delta).$$

According to (37) in connexion with (45), (46), the degree in δ of $a^{(2h,3h-2)}_{2(h-\lambda)}(\delta)$ is $2h$. The highest term, $a_0\delta^{2h}$, is contained in the term corresponding to $\rho = 0$ in (36). Inserting in (36) the values from (44), (45) and (46) and putting in the sum $\rho = 0$, we have

$$2^{-2h+\lambda+1}\frac{(2h-\delta)^{[2\lambda]}}{\lambda!}d_{2h-\delta,0}\,d^{(4h-\lambda-2\delta)}_{2\delta,h-\lambda}.$$

Thus, in virtue of (16) and (II), para. 10,

$$a_0 = 2^{-2h+\lambda+1}\frac{1}{\lambda!}(-1)^{h-\lambda}\frac{2^{2h-2\lambda}}{2^{h-\lambda}(h-\lambda)!} = \frac{(-1)^{h-\lambda}}{2^{h-1}h!}\binom{h}{\lambda}.$$

According to (14), $$K^{(2h,3h-2)}_{2(h-\lambda),\lambda} = \frac{(-1)^{h-\lambda}(2h)!}{2^{h-1}h!}\binom{h}{\lambda}.$$

Inserting this in (47) we have

$$R_h = \frac{(2h)!}{2^{h-1}h!}\sum_{\lambda=0}^{h}(-1)^{h-\lambda}\binom{h}{\lambda}p^{2(h-\lambda)}k^{\lambda} = \frac{(2h)!}{2^{h-1}h!}(k-p^2)^h.$$

The highest term of $\mu_{2h}(p')$ is thus

$$2^h\frac{(2h)!}{h!}(k-p^2)^h n^{-h},$$

that of $\mu_2(p')$ is $4(k-p^2)n^{-1}$, and hence

$$\alpha_{2h}(p') = \frac{\mu_{2h}(p')}{\mu_2^h(p')} \xrightarrow[n]{} \frac{(2h)!}{2^h h!} \quad \text{if} \quad k-p^2>0. \tag{48}$$

From (42) and (48) it follows according to the Second Limit Theorem that the distribution of p' tends to normality as $n\to\infty$, provided that the marginal distributions are continuous and $k-p^2>0$.

The condition $k-p^2>0$ is fulfilled if the population is distributed normally. For, comparing Esscher's formula (5) with (13), we find, since $\operatorname{var}(\tau) = 4\operatorname{var}(p')$,

$$k-p^2 = \frac{1}{36} - \left\{\frac{1}{\pi}\sin^{-1}\left(\frac{r}{2}\right)\right\}^2.$$

The right-hand side is positive if $|r|<1$.

V. THE VARIANCE OF p' IN THE CASE OF A FINITE POPULATION

19. Consider a sample of n drawn from a finite population of N in which all values of x and all values of y are different. For the sample probabilities p', k', $(p^2)'$, ... we write now

$$p^{(n)}, \quad k^{(n)}, \quad (p^2)^{(n)}, \quad \ldots,$$

and for the corresponding population probabilities

$$p^{(N)}, \quad k^{(N)}, \quad (p^2)^{(N)}, \quad \ldots.$$

Equation (9) remains valid and may be written as follows:

$$Ew^{(n)} = w^{(N)}. \tag{49}$$

In particular, $$Ep^{(n)} = p^{(N)}.$$

The essential difference between this case and the case $N=\infty$ considered above is that the composite probabilities such as $(p^2)^{(N)}$ or $(pk)^{(N)}$ are not equal to $(p^{(N)})^2$ or $p^{(N)}k^{(N)}$. For

instance, $(p^{(N)})^2$ is evidently the same function of $p^{(N)}$, $k^{(N)}$, $(p^2)^{(N)}$ and N as $(p')^2$ is of p', k', $(p^2)'$ and n. Thus we have, replacing n by N in (10),

$$(p^{(N)})^2 = \frac{2}{N^{[2]}}p^{(N)} + \frac{4(N-2)}{N^{[2]}}k^{(N)} + \frac{(N-2)^{[2]}}{N^{[2]}}(p^2)^{(N)}, \tag{50}$$

and hence

$$(p^2)^{(N)} = \frac{N^{[2]}}{(N-2)^{[2]}}(p^{(N)})^2 - \frac{4}{N-3}k^{(N)} - \frac{2}{(N-2)^{[2]}}p^{(N)}. \tag{51}$$

On the other hand, from (10) and (49)

$$E(p^{(n)})^2 = \frac{2}{n^{[2]}}p^{(N)} + \frac{4(n-2)}{n^{[2]}}k^{(N)} + \frac{(n-2)^{[2]}}{n^{[2]}}(p^2)^{(N)}, \tag{52}$$

which is the equivalent of equation (11).

On subtracting (50) from (52) we find

$$\begin{aligned}\operatorname{var}(p^{(n)}) = \frac{2(N-n)}{n(n-1)N(N-1)}\{(N+n-1)p^{(N)} + 2[Nn-2(N+n-1)]k^{(N)} \\ - [2Nn-3(N+n-1)](p^2)^{(N)}\}.\end{aligned} \tag{53}$$

Substituting for $(p^2)^{(N)}$, the expression in (51), we obtain

$$\begin{aligned}\operatorname{var}(p^{(n)}) = \frac{2(N-n)}{n(n-1)(N-2)(N-3)}\{(N+n-5)p^{(N)}(1-p^{(N)}) \\ + 2(n-2)(N-2)(k^{(N)} - p^{(N)2})\},\end{aligned} \tag{54}$$

or

$$\binom{n}{2}\operatorname{var}(p^{(n)}) = \left(1 - \frac{(n-2)(n-3)}{(N-2)(N-3)}\right)p^{(N)}(1-p^{(N)}) + 2(n-2)\left(1 - \frac{n-3}{N-3}\right)(k^{(N)} - p^{(N)2}). \tag{55}$$

For $N \to \infty$, (55) becomes the same as (13).

VI. A sample estimate of var (p')

20. In the case of an infinite population, let

$$\binom{n}{2}\operatorname{var}'(p') = p' + 2(n-2)k' - (2n-3)(p^2)'. \tag{56}$$

Then, in virtue of (9) and (12),

$$E\operatorname{var}'(p') = \operatorname{var}(p').$$

On inserting in (56) for $(p^2)'$ the expression obtained from (10), we find

$$\binom{n-2}{2}\operatorname{var}'(p') = p'(1-p') + 2(n-2)(k' - (p')^2),$$

or

$$\operatorname{var}'(p') = \frac{2}{(n-2)(n-3)}p'q' + \frac{4}{n-3}(k' - (p')^2). \tag{57}$$

By analogy,

$$\operatorname{var}'(q') = \frac{2}{(n-2)(n-3)}p'q' + \frac{4}{n-3}(l' - (q')^2), \tag{58}$$

where l' is the probability that among three members A, B, C drawn from the sample without replacement, one, say A, is discordant with the other two.

In the case of a finite population of the type considered in para. 19, we define in a similar way a statistic $\operatorname{var}^{(n)}(p^{(n)})$ such that

$$E\operatorname{var}^{(n)}(p^{(n)}) = \operatorname{var}(p^{(n)}).$$

We find

$$\operatorname{var}^{(n)}(p^{(n)}) = \frac{2(N-n)}{N(N-1)(n-2)(n-3)}\{(N+n-5)p^{(n)}(1-p^{(n)}) + 2(n-2)(N-2)(k^{(n)}-p^{(n)2})\}. \tag{59}$$

A comparison between (59) and (54) shows that $\operatorname{var}^{(n)}(p^{(n)})$ is obtained from $\operatorname{var}(p^{(n)})$ by interchanging n and N and taking the opposite sign.

21. Let g_ν and h_ν be the numbers of sample members concordant and discordant with $A_\nu = (x_\nu, y_\nu)$ $(\nu = 1, \ldots, n)$. The probability of drawing first the member A_ν, and then, without replacing it, a member concordant with A_ν is $\frac{1}{n}\frac{g_\nu}{n-1}$. The probability of drawing, without replacement, first A_ν and then two other members concordant with A_ν is $\frac{1}{n}\frac{g_\nu(g_\nu-1)}{(n-1)(n-2)}$. Hence

$$p' = \frac{\Sigma g_\nu}{n(n-1)}, \quad k' = \frac{\Sigma g_\nu^2 - \Sigma g_\nu}{n(n-1)(n-2)}. \tag{60}$$

Similarly,

$$q' = \frac{\Sigma h_\nu}{n(n-1)}, \quad l' = \frac{\Sigma h_\nu^2 - \Sigma h_\nu}{n(n-1)(n-2)}. \tag{61}$$

If only the value of p' or q' is required, the use of (6) may be more expedient than that of (60) or (61). If, however, the variance, and hence k' or l', is wanted, the calculation by means of the numbers g_ν and h_ν (whose sums are twice the numbers of inversions I, J) according to (60) or (61) is to be preferred.

If $p' > \frac{1}{2}$, it is more convenient to calculate q' and l' from (61); if $p' < \frac{1}{2}$, the calculation of p' and k' by (60) is more rapid. In many cases one can see directly from the given data whether the concordant or the discordant pairs prevail, before actually calculating p' or q'.

Since $p' + q' = 1$, we have

$$\operatorname{var}(p') = \operatorname{var}(q'),$$

and also, in the case of a finite population,

$$\operatorname{var}(p^{(n)}) = \operatorname{var}(q^{(n)}).$$

If we write down the equation for $\operatorname{var}(q^{(n)})$ analogous to (55) and subtract it from (55), we have

$$k^{(N)} - p^{(N)2} - l^{(N)} + q^{(N)2} = 0,$$

or

$$k^{(N)} - l^{(N)} = p^{(N)2} - q^{(N)2} = p^{(N)} - q^{(N)}.$$

Substituting n for N, we have

$$k' - l' = p' - q' = \tau.$$

Comparing this with (57) and (58) we see that

$$\operatorname{var}'(p') = \operatorname{var}'(q').$$

REFERENCES

ESSCHER, F. (1924). On a method of determining correlation from the ranks of the variates. *Skand. Aktuar.* **7**, 201–19.

JORDAN, CH. (1927). *Statistique mathématique.* Paris: Gauthiers-Villars.

KENDALL, M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**, 81–93.

LINDEBERG, J. W. (1926). Ueber die Korrelation. *Den VI skandinaviske Matematikerkongres i København*, 31 August–4 September 1925, pp. 437–46. København: J. Gjellerup.

ADDENDUM

On p. 184 above, I quoted J. W. Lindeberg as having given the formula for the variance of the probability of concordance p' without proof. I was not aware then that a proof of this formula, as well as that of the corresponding expression for a finite population (equation (54) of my paper), is contained in another paper by Lindeberg, 'Some remarks on the mean error of the percentage of correlation,' *Nordic Statistical Journal*, **1**, 137–41 (1929).

THE SIGNIFICANCE OF RANK CORRELATIONS WHERE PARENTAL CORRELATION EXISTS

By H. E. Daniels (*Wool Industries Research Association*)
and M. G. Kendall

1. All the known tests of significance of rank correlation coefficients are based on distributions from a population in which each possible ranking occurs equally frequently, i.e. the null case where no parental correlation exists. We may then say of any particular coefficient whether it is significant in the sense that it cannot have arisen with any acceptable probability from an uncorrelated population. No tests are known in the case where parental correlation exists, and we have not seen the point discussed except in reference to the replacement of rank correlations by grade or product-moment correlations. Thus, for example, if two rank correlation coefficients are both found to be significant there has hitherto been no exact method of deciding whether their difference is significant. In this paper we consider the problem of determining confidence intervals for a rank correlation when the parent is correlated and develop a test of significance for the difference of two correlations.

2. In testing an ordinary product-moment correlation the problem is enormously simplified by the assumption that the population is normal, or the further assumption that normal theory holds good even when the parent deviates only moderately from normality. Apart from means and variances the population is then completely specified by the single parent parameter ρ and, as is well known, the sample distribution of the estimator depends only on ρ and the sample number n.

In ranking theory this position no longer obtains. No assumption can in general be made about the form of the parent distribution and, in particular, the parent correlation does not completely specify the problem. The usual type of variate theory cannot, therefore, be expected to meet the requirements.

3. A satisfactory approach to the problem can, however, be made if the rank correlation is measured by the coefficient known as τ (Kendall, 1943, chap. 16). We shall then show that, for large samples at any rate, the problem admits of a solution.

Let the population consist of N members. They may be imagined as laid out in the natural order $1, 2, \ldots, N$ according to the first variate. The rankings according to the second variate are then some permutation of the numbers 1 to N, and this second array of ranks is all we need write down in particular cases. It determines the rank correlation τ. Now suppose we choose a sample of n in one of the $\binom{N}{n}$ possible ways. This sample will, so far as the first variate is concerned, be in the natural order, and the ranks according to the second variate permit of the calculation of a sample correlation t. For all possible samples and any given arrangement of the parent members there will be a distribution of $\binom{N}{n}$ values of t.

4. The sample value of t is an unbiased estimator of τ; that is to say, the mean value of t in all possible samples is τ. For consider the $\binom{N}{n}$ samples of n. Any particular pair of members will occur in $\binom{N-2}{n-2}$ samples, that is, all pairs occur equally frequently in the totality of all samples. In calculating t we assign to any pair $+1$ if its members are in the right order and -1 in the contrary case. Thus the total of the score for all samples is $\binom{N-2}{n-2}$ times the score for the population. To obtain t we divide the score for any sample by $\frac{1}{2}n(n-1)$, and to obtain τ we divide the population score by $\frac{1}{2}N(N-1)$. Hence if Σ is the score for the population, the mean value (expectation) of t is

$$E(t) = \frac{\binom{N-2}{n-2}\Sigma}{\frac{1}{2}n(n-1)\binom{N}{n}} = \frac{\Sigma}{\frac{1}{2}N(N-1)} = \tau. \tag{4·1}$$

5. Unfortunately, it is not true that higher moments of t depend only on τ. A single example will illustrate the point. Consider the ranking of 9:

5 2 3 1 6 7 8 9 4.

If the $84 = \binom{9}{3}$ possible samples of three are written down and t evaluated for each, the distribution of S (the number of positive pairs) is found to be as follows:

Values of S	Frequency
0	2
1	15
2	34
3	33
Total	84

The mean of this distribution is $182/84 = 13/6$, and since

$$t = \frac{2S}{\frac{1}{2}n(n-1)} - 1,$$

the mean value of t is $(26/18) - 1 = 0·44$. The value of S for the parent ranking is 26 and hence $\tau = (52/36) - 1 = 0·44$, verifying equation (4·1). The ranking

1 2 5 9 3 6 7 8 4

also has $\tau = 0·44$, but the distribution of S in samples of three is now:

Values of S	Frequency
0	3
1	16
2	29
3	36
Total	84

The second moment of this distribution is 5·429, against 5·333 for the first distribution, the variances being 0·734 against 0·639.

6. Thus for any parent with given τ there is in general more than one sampling distribution of t according to the arrangement of the parent ranks. In short, as mentioned above, the parameter τ does not completely specify the sampling distribution and in asking the question: What is the standard error of t? we are seeking for an answer which does not exist.

It will be shown, however, that for any given parent ranking the distribution of t tends to normality with increasing n. The sampling properties of t can therefore be specified to a first approximation by its first and second moments only, when the samples are not too small. Further, it will be proved that for given τ the variance of t cannot exceed a certain function of τ and n whatever the parent ranking. From a knowledge of t and n only, it is thus possible to set outer bounds to confidence intervals for τ provided n is large enough for the normal approximation to hold. The limits obtained in this way are sometimes rather wide, and an alternative procedure is to estimate the true variance of t directly from the sample itself according to a formula given below. This avoids the loss of efficiency consequent on using an upper limit to the variance, but it is not known how large a sample is required for the error of estimation to be tolerable.

7. The development of the theory is facilitated if we introduce at the present stage a notation similar to that used by Daniels (1944). The ith and jth ranks corresponding to the second variate are together assigned a score a_{ij} which takes the value $+1$ if the members are in the correct order, -1 if in the wrong order, and a_{ii} is defined to be zero. The ranks for the first variate are similarly assigned scores b_{ij}, but as the members have been taken in the correct order for this variate, the scores are simply $b_{ij} = \pm 1$, $i \lessgtr j$; $b_{ii} = 0$. Next we define $c_{ij} = a_{ij}b_{ij}$, so that $c_{ij} = \pm 1$, according to whether the ranks for the two variates agree or differ in order, and $c_{ii} = 0$. In this notation

$$\tau = c/N(N-1),$$

where $c = \Sigma c_{ij}$, i and j both being summed from 1 to N.

When the sample of n pairs is selected at random from the parent N and its coefficient t is calculated, the values of c_{ij} for the members of the sample remain the same as in the population. This fact makes τ much more suitable for the present problem than the Spearman coefficient ρ whose associated scores do not possess the same property. The sample rank correlation is then

$$t = c^{(n)}/n(n-1),$$

where $c^{(n)} = \Sigma^{(n)} c_{ij}$ and $\Sigma^{(n)}$ denotes summation only over those values of i and j occurring in the sample.

8. It has already been proved that $E(t) = \tau$. To find the variance of t we require $E(t^2)$, so consider

$$\sum_n [c^{(n)}]^2 = \sum_n \Sigma^{(n)} c_{ij} c_{kl},$$

$\sum\limits_n$ denoting summation over all selections of the sample of n from the finite parent population of N members. Let us enumerate the number of ways in which $c_{ij}c_{kl}$ and similar products with 'tied' suffixes, such as $c_{ij}c_{il}$, occur in the sum.

(i) When i, j, k, l are all different the term $c_{ij}c_{kl}$ may occur with $\binom{N-4}{n-4}$ selections of the remaining members of the sample and the contribution of such terms to $\sum\limits_n$ is $\binom{N-4}{n-4}\Sigma' c_{ij}c_{kl}$, Σ' meaning summation over all unequal values of i, j, k, l from 1 to N.

(ii) The term $c_{ij}c_{il}$ similarly occurs in $\binom{N-3}{n-3}$ ways and there are four ways of tying one suffix, each of which gives the same contribution to Σ since c_{ij} is symmetrical. The total contribution of such terms to $\sum\limits_n$ is therefore $4\binom{N-3}{n-3}\overset{n}{\Sigma'} c_{ij}c_{il}$.

(iii) Terms like $c_{ij}c_{ij}$ similarly contribute $2\binom{N-2}{n-2}\Sigma' c_{ij}c_{ij}$ to $\sum\limits_u$, and all other terms are zero since $c_{ii} = 0$. Hence

$$\sum_n [c^{(n)}]^2 = \binom{N-4}{n-4}\Sigma' c_{ij}c_{kl} + 4\binom{N-3}{n-3}\Sigma' c_{ij}c_{ik} + 2\binom{N-2}{n-2}\Sigma' c_{ij}c_{ij}.$$

Expressing the Σ''s in terms of the corresponding Σ's and dividing out by $\binom{N}{n}$ we obtain

$$E[c^{(n)}]^2 = \frac{n^{(4)}}{N^{(4)}}(\Sigma c_{ij}c_{kl} - 4\Sigma c_{ij}c_{il} + 2\Sigma c_{ij}c_{ij}) + \frac{4n^{(3)}}{N^{(3)}}(\Sigma c_{ij}c_{il} - \Sigma c_{ij}c_{ij}) + \frac{2n^{(2)}}{N^{(2)}}\Sigma c_{ij}c_{ij},$$

where $n^{(r)} = n(n-1)\dots(n-r+1)$. Since $\Sigma c_{ij}c_{ij} = N(N-1)$ and $\Sigma c_{ij}c_{kl} = c^2$, the variance of t for given τ and n is seen to depend on the value of $\Sigma c_{ij}c_{ik} = \Sigma c_i^2$, where $c_i = \sum\limits_{j=1}^{N} c_{ij}$.

Let N become large. The quantities c and Σc_i^2 are respectively $O(N^2)$ and $O(N^3)$, so if we introduce $\tau_i = c_i/N$ the value of $E(t^2)$ for large N becomes

$$E(t^2) \sim \frac{(n-2)(n-3)}{n(n-1)}\tau^2 + \frac{4(n-2)}{n(n-1)}\frac{\Sigma\tau_i^2}{N} + \frac{2}{n(n-1)},$$

and hence in the limit the variance of t is

$$\operatorname{var} t = \frac{4(n-2)}{n(n-1)}\operatorname{var}\tau_i + \frac{2}{n(n-1)}(1-\tau^2). \tag{8·1}$$

9. The variance of t satisfies the inequality

$$\operatorname{var} t \leqslant \frac{2}{n}(1-\tau^2), \tag{9·1}$$

whatever the parent ranking. Moreover, though the limit may not be attained in any particular parent ranking, reasons are given in the Appendix for expecting that it cannot be substantially improved upon. The proof is as follows.

Reverting to a finite parent population of N members, we first seek a maximum for Σc_i^2. In terms of the original scores, $c_{ij} = a_{ij}b_{ij}$. Keeping $b_{ij} = \pm 1$, $i \lessgtr j$, $b_{ii} = 0$, as before, allow the a_{ij}'s to assume *any* values subject to the conditions

$$\Sigma a_{ij}^2 = N(N-1), \quad \Sigma a_{ij}b_{ij} = c = N(N-1)\tau.$$

The stationary values of Σc_i^2 occur when the a_{ij}'s satisfy the equations

$$b_{ij}(c_i + c_j) - \lambda a_{ij} - \mu b_{ij} = 0,$$

which give, on multiplying by b_{ij} and summing j,

$$c_i = \frac{\mu(N-1)-c}{(N-2-\lambda)}.$$

Thus, unless the c_i's are all to be equal, in which case Σc_i^2 is a minimum, λ and μ must take the values

$$\lambda = N-2, \quad \mu = c/(N-1),$$

and since

$$2\Sigma c_i^2 - \lambda N(N-1) - \mu c = 0,$$

it follows that Σc_i^2 cannot exceed $\frac{1}{2}N(N-1)(N-2)+\frac{1}{2}c^2/(N-1)$. Allowing N to become large, this implies

$$\Sigma \tau_i^2/N \leqslant \tfrac{1}{2}(1+\tau^2).$$

Hence

$$\operatorname{var} \tau_i \leqslant \tfrac{1}{2}(1-\tau^2),$$

and so from equation (8·1)

$$\operatorname{var} t \leqslant \frac{2}{n}(1-\tau^2). \tag{9·1}$$

10. Assuming that the sample is large enough for the distribution of t to be normal, the roots τ_1, τ_2 of the equation

$$t-\tau = x\sqrt{\left[\frac{2}{n}(1-\tau^2)\right]}, \tag{10·1}$$

i.e.

$$\tau = \frac{t \pm x\sqrt{\frac{2}{n}}\sqrt{\left(1+\frac{2x^2}{n}-t^2\right)}}{\left(1+\frac{2x^2}{n}\right)}, \tag{10·2}$$

provide confidence limits to τ when t is known, x being the standardized normal deviate corresponding to a given probability of P %. These confidence limits are of course maxima, in the sense that we shall be wrong in *at most* P % of the cases in asserting τ to lie between the calculated limits.

In our proof of the tendency of t to normality it will be necessary to neglect terms of order $n^{-\frac{1}{2}}$, and the sample may have to be rather large for such terms to be small, unless τ itself is small.

The form of equation (9·1) suggests using

$$w = \sin^{-1} t$$

instead of t. To the same order of approximation we can take w as having a normal distribution with mean $\omega = \sin^{-1}\tau$ and standard error not exceeding $\sqrt{(2/n)}$, which is independent of τ. This form is more convenient for assigning confidence limits to t, and for testing the significance of the difference between t_1 and t_2 (whose standard error cannot exceed $\sqrt{[2(1/n_1+1/n_2)]}$), but we have not been able to discover whether the transformation brings the distribution nearer to normality.

11. We now prove that the distribution of t tends for large n to normality whatever the parent ranking, provided that $|\tau|$ is not near unity.

Write $g_{ij} = c_{ij} - c/N^2$ so that $\Sigma g_{ij} = 0$, $g_{ij} = g_{ji}$ and $g_{ii} = -c/N^2 = -(N-1)\tau/N$. The rth moment of $c^{(n)}$ about its mean value is $E[\Sigma^{(n)} g_{ij}]^r$, so consider

$$\sum_n [\Sigma^{(n)} g_{ij}]^r = \sum_n \Sigma^{(n)} g_{ij} g_{kl} g_{uv} \cdots,$$

the summation $\sum_n$ being over all possible sample selections.

The argument used by Daniels (1944) to show that in the null case the distribution of rank correlation in large samples tends to normality can be applied with little modification to the present problem. The proof is therefore sketched here without much detail.

Two essential conditions to be satisfied are that $\Sigma g_{ij} = 0$, which is true by definition, and $\Sigma g_{ij} g_{ik} = O(N^3)$, which is true only if $1-\tau^2 = O(1)$, so that the tendency to normality may be expected to break down for high correlations.

The sum $\underset{n}{\Sigma}$ is evaluated as in § 8 by counting the number of ways in which terms like $g_{ij} g_{kl} g_{uv} \ldots$, and similar terms with tied suffixes, occur. In this way it is expressed as a linear combination of $\Sigma' g_{ij} g_{kl} g_{uv} \ldots$, etc. Every such Σ' is replaceable by the corresponding Σ together with terms containing more tied suffixes which are of lower order in N since they involve fewer summations from 1 to N.

12. First consider the even moments with $r = 2m$. Terms containing more than $3m$ different suffixes must vanish, since in such cases it is impossible to avoid at least one g_{ij} with two free suffixes, and $\Sigma g_{ij} = 0$. For the same reason the only non-vanishing terms with $3m$ different suffixes are those containing expressions like

$$\Sigma g_{ij} g_{ik} g_{lu} g_{lv} g_{pq} g_{pr} \ldots = (\Sigma g_{ij} g_{ik})^m,$$

and terms with fewer different suffixes are of correspondingly lower order in N.

With $3m$ suffixes assigned there are $\binom{N-3m}{n-3m}$ ways of selecting the remaining $n-3m$ members of the sample, and the suffixes can be tied in $\frac{(2m)!\,2^{2m}}{m!\,(2!)^m}$ ways to give the same result. Dividing out by $\binom{N}{n}$ and noting that $\binom{N-3m}{n-3m}\Big/\binom{N}{n} \sim n^{3m}/N^{3m}$ when both N and n are large, the contribution of such terms to μ_{2m}, the $2m$th moment of $c^{(n)}$ about its mean, is found to be

$$\frac{n^{3m}}{N^{3m}} \frac{(2m)!}{m!} 2^m (\Sigma g_{ij} g_{ik})^m,$$

which is of order n^{3m}. Moreover, by the same argument, terms with $f < 3m$ different suffixes add contributions of order n^f which may be neglected.

Hence
$$\mu_{2m} \sim \frac{n^{3m}}{N^{3m}} \frac{(2m)!}{m!} 2^m (\Sigma g_{ij} g_{ik})^m,$$

the neglected terms being relatively $O(n^{-1})$.

13. For the odd moments let $r = 2m+1$. Similar considerations show that the non-vanishing terms of $\underset{n}{\Sigma}$ cannot have more than $3m+1$ different suffixes, and μ_{2m+1} is therefore of order n^{3m+1}.

Then since $c^{(n)}/n^{\frac{3}{2}}$ has even moments of unit order and odd moments of order $n^{-\frac{1}{2}}$, the odd moments may be neglected to that order. We conclude that $c^{(n)}$ is distributed normally for large n with variance

$$\frac{4n^3}{N^3} \Sigma g_{ij} g_{ik} = 4n^3 \operatorname{var} \tau_i,$$

and t is similarly normal with variance $(4/n) \operatorname{var} \tau_i$.

14. The fact that terms of order $n^{-\frac{1}{2}}$ have to be neglected suggests that the normal approximation only holds good for fairly large samples. This is not surprising since one would expect skewness to be an important property of the distribution of t when τ is not zero, if only for the reason that $|t|$ can never exceed unity. It seems worth while to examine the odd moments in more detail.

The dominant term of the $(2m+1)$th moment has $3m+1$ different suffixes, which can occur as
$$\Sigma g_{ij} g_{ik} g_{il} (\Sigma g_{uv} g_{uw})^{m-1} \quad \text{or} \quad \Sigma g_{ij} g_{ik} g_{jl} (\Sigma g_{uv} g_{uw})^{m-1}.$$
Both can be obtained in
$$\frac{(2m+1)!\, 2^{2(m-1)} 2^3}{(2!)^{m-1} 3!(m-1)!} = \frac{(2m+1)!\, 2^{m+2}}{3!(m-1)!}$$
distinct ways, and there are $\binom{N-3m-1}{n-3m-1}$ ways of selecting the sample with $3m+1$ suffixes assigned. The $(2m+1)$th moment of $c^{(n)}$ about its mean is therefore
$$\mu_{2m+1} \sim \frac{n^{3m+1}}{N^{3m+1}} \frac{(2m+1)!\, 2^{m+2}}{3!(m-1)!} [\Sigma g_{ij} g_{ik} g_{il} + \Sigma g_{ij} g_{ik} g_{jl}] (\Sigma g_{ij} g_{ik})^{m-1},$$
ignoring terms of relative order $O(n^{-1})$. The corresponding moment of t is obtained to the same order on dividing by n^{4m+2}; it depends only on $\operatorname{var} t$ and $\mu_3(t)$, where
$$\mu_3(t) \sim \frac{8}{n^2 N^4} [\Sigma g_{ij} g_{ik} g_{il} + \Sigma g_{ij} g_{ik} g_{jl}] = \frac{4}{n^2} \frac{\Sigma g_{ij}(g_i + g_j)^2}{N^4},$$
where $g_i = \sum_{j=1}^{N} g_{ij}$. The distribution of t is thus specified to $O(n^{-1})$ by its first three moments. The moment-generating function of the distribution of t in standard measure is
$$M(z) = \left(1 + \frac{\gamma_1}{3!} z^3\right) e^{-\frac{1}{2}z^2} [1 + O(n^{-1})],$$
where
$$\gamma_1 = \mu_3(t)/(\operatorname{var} t)^{\frac{3}{2}} = O(n^{-\frac{1}{2}}),$$
and the frequency distribution of $x = (t-\tau)/\sqrt{(\operatorname{var} t)}$ is*
$$f(x) = \left(1 - \frac{\gamma_1}{3!} \frac{d^3}{dx^3}\right) \frac{e^{-\frac{1}{2}x^2}}{\sqrt{(2\pi)}} [1 + O(n^{-1})]. \tag{14·1}$$

15. The effect of the γ_1 term in modifying the confidence limits based on normal theory can be seen in the following way. Let ξ be the normal deviate whose chance of being exceeded is $P(\xi)$. The chance of x exceeding ξ is, from (14·1),
$$F(\xi) = P(\xi) + \frac{\gamma_1}{6} (\xi^2 - 1) \frac{e^{-\frac{1}{2}\xi^2}}{\sqrt{(2\pi)}}.$$
If X is the correct limit such that $F(X) = P(\xi)$, it is readily proved by successive approximation that the formula
$$X = \xi + \frac{\gamma_1}{6} (\xi^2 - 1) \tag{15·1}$$
gives the appropriate value of X to $O(n^{-1})$. For example, the 5 and 1 % limits are respectively $\pm 1{\cdot}96 + 0{\cdot}474\gamma_1$ and $\pm 2{\cdot}58 + 0{\cdot}941\gamma_1$.

16. In practice the value of $\operatorname{var} t$ has to be estimated from the sample, and although its standard error can be shown to be $O(n^{-\frac{1}{2}})$ by the kind of argument already used, it is not known how large the sample has to be before the error in estimating the variance can be safely ignored. It is best to use the unbiased formula
$$\operatorname{var} t = \frac{1}{n(n-1)(n-2)(n-3)} \left\{4\Sigma c_i^2 - \frac{2(2n-3)}{n(n-1)} c^2 - 2n(n-1)\right\} \tag{16·1}$$
(which is easily proved) in calculating $\operatorname{var} t$ from the sample, especially if the standard error of the mean value of t from a number of small samples is required.

* Note that the approximation error in $f(x)$ is relatively $O(n^{-1})$, a stronger result than would be obtained from a Gram-Charlier approximation based on the first three moments only.

As the term in γ_1 is a small correction it is perhaps sufficient in moderate samples to take

$$G = \tfrac{1}{2}\Sigma g_{ij}(g_i+g_j)^2 = \tfrac{1}{2}\Sigma c_{ij}(c_i+c_j)^2 - \frac{5c\Sigma c_i^2}{n} + \frac{3c^2}{n^3}, \tag{16·2}$$

and
$$\mu_3(t) = \frac{8}{n^6}G, \quad \gamma_1 = \mu_3(t)/(\mathrm{var}\, t)^{\frac{3}{2}}, \tag{16·3}$$

where the first term in G is the sum of $c_{ij}(c_i+c_j)^2$ over all values of $i>j$. The unbiased formula for $\mu_3(t)$ involves some rather tedious computation.

17. To illustrate the methods of the paper we consider an actual example.

A set of thirty wool samples were visually graded in order of fibre fineness by three assessors. The mean fibre diameter for each wool sample was also determined by direct measurement. Table 1 shows the measured order (M) compared with that of the three assessors (A, B, C), in ascending order of experience.

Table 1

M	A	B	C	M	A	B	C
1	5	2	1	16	12	14	16
2	4	5	2	17	10	18	15
3	9	6	6	18	30	21	25
4	3	1	3	19	22	26	24
5	6	7	4	20	16	22	19
6	2	4	5	21	21	16	18
7	15	19	10	22	29	20	23
8	18	3	12	23	28	25	22
9	8	8	7	24	19	27	26
10	11	9	8	25	23	28	21
11	17	13	9	26	20	23	27
12	13	10	11	27	7	24	20
13	24	17	17	28	26	29	28
14	14	12	14	29	27	15	30
15	1	11	13	30	25	30	29

The method of working will be seen from the c_{ij} matrix for the MA correlation shown in Table 2.

The correlations of the assessors' orders with the measured order are found to be

$$t_A = 0·490, \quad t_B = 0·724, \quad t_C = 0·816.$$

(i) Consider first the maximum confidence limits given by (10·2). The 5 % limits are

$$-0·02 < t_A < 0·80, \quad 0·23 < t_B < 0·92, \quad 0·34 < t_C < 0·96.$$

Again, using the transformation $w = \sin^{-1} t$, the 5 % limits are

$$0·01 < t_A < 0·85, \quad 0·30 < t_B < 0·97, \quad 0·45 < t_C < 0·99.$$

The values of w are $\quad w_A = 0·512, \quad w_B = 0·810, \quad w_C = 0·954.$

The greatest difference is 0·442, and the upper limit to its standard error is $\sqrt{(4/n)} = 0·365$, so on these grounds the difference between A and C would not be judged significant.

The 5 % limits are very wide, and the lack of significance is disappointing since C was known to be an expert appraiser while A is relatively inexperienced, and one would have expected an obvious difference between them.

(ii) The variances estimated from the unbiased formula (16·1) are

$$\operatorname{var} t_A = 0{\cdot}006630, \quad \operatorname{var} t_B = 0{\cdot}005067, \quad \operatorname{var} t_C = 0{\cdot}002198.$$

The estimated standard errors are therefore

$$s_A = 0{\cdot}081, \quad s_B = 0{\cdot}071, \quad s_C = 0{\cdot}047.$$

The 5 % confidence limits, assuming normality, are

$$0{\cdot}33 < t_A < 0{\cdot}65, \quad 0{\cdot}58 < t_B < 0{\cdot}86, \quad 0{\cdot}72 < t_C < 0{\cdot}91.$$

Table 2

															c_{ij}															c_i
0	−	+	−	+	−	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	21
−	0	+	−	+	−	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	21
+	+	0	−	−	−	+	+	−	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	17
−	−	−	0	+	−	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	19
+	+	−	+	0	−	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	23
−	−	−	−	−	0	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	17
+	+	+	+	+	+	0	+	−	−	+	−	+	−	−	−	−	+	+	+	+	+	+	+	+	+	−	+	+	+	13
+	+	+	+	+	+	+	0	−	−	−	−	+	−	−	−	−	+	+	−	+	+	+	+	+	+	−	+	+	+	9
+	+	−	+	+	+	−	−	0	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	19
+	+	+	+	+	+	−	−	+	0	+	+	+	+	−	+	−	+	+	+	+	+	+	+	+	+	−	+	+	+	19
+	+	+	+	+	+	+	−	+	+	0	−	+	−	−	−	−	+	+	−	+	+	+	+	+	+	−	+	+	+	13
+	+	+	+	+	+	−	−	+	+	−	0	+	+	−	−	−	+	+	+	+	+	+	+	+	+	−	+	+	+	15
+	+	+	+	+	+	+	+	+	+	+	+	0	−	−	−	−	+	−	−	−	+	+	−	−	−	−	+	+	+	7
+	+	+	+	+	+	−	−	+	+	−	+	−	0	−	−	−	+	+	+	+	+	+	+	+	+	−	+	+	+	13
−	−	−	−	−	−	−	−	−	−	−	−	−	−	0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	1
+	+	+	+	+	+	−	−	+	+	−	−	−	−	+	0	−	+	+	+	+	+	+	+	+	+	−	+	+	+	13
+	+	+	+	+	+	−	−	+	−	−	−	−	−	+	−	0	+	+	+	+	+	+	+	+	+	−	+	+	+	11
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0	−	−	−	−	−	−	−	−	−	−	−	−	5
+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	−	0	−	−	+	+	−	+	−	−	+	+	+	15
+	+	+	+	+	+	+	−	+	+	−	+	−	+	+	+	+	−	−	0	+	+	+	+	+	+	−	+	+	+	17
+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	−	−	+	0	+	+	−	+	+	−	+	+	+	19
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	0	−	−	−	−	−	−	−	−	11
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	−	0	−	−	−	−	−	−	−	11
+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	−	−	+	−	−	−	0	+	+	−	+	+	+	15
+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	−	+	+	+	−	−	+	0	−	−	+	+	+	17
+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	−	−	+	+	−	−	+	−	0	−	+	+	+	15
+	+	−	+	+	+	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	0	+	+	+	−11
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	−	−	+	+	+	+	0	+	−	21
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	−	−	+	+	+	+	+	0	−	21
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	−	−	+	+	+	+	−	−	0	19

$c =$ 426

$n =$ 30

$\Sigma c_i^2 = 7470$

Moreover, we should judge A and C to be significantly different at the 1 % level, and A and B at the 5 % level. How far these conclusions are valid depends, of course, on the accuracy of the variance estimates, but the conclusions seem to agree with what might have been expected from prior knowledge of the assessors' capabilities.

(iii) The values of γ_1 calculated from (16·2) and (16·3) are

$$\gamma_1(A) = -0{\cdot}32, \quad \gamma_1(B) = -0{\cdot}35, \quad \gamma_1(C) = -0{\cdot}38.$$

The distributions would not appear to be very skew, and the distribution of the difference of two t's is probably nearly normal. The adjusted 5 % limits are, from (15·1),

$$0{\cdot}32 < t_A < 0{\cdot}64, \quad 0{\cdot}57 < t_B < 0{\cdot}85, \quad 0{\cdot}72 < t_C < 0{\cdot}90.$$

APPENDIX

1. The question arises whether a particular parental form exists for which the variance of t assumes the upper limit $2(1-\tau^2)/n$. We surmise, though we cannot prove, that the maximum possible variance is attained when the parent ranking has a 'canonical' form obtained in the following way. Consider again the ranking

$$5\quad 2\quad 3\quad 1\quad 6\quad 7\quad 8\quad 9\quad 4.$$

The number of positive pairs S is 26, so that $t = 0{\cdot}44$. Let us transform this so as to bring the 1 to the beginning of the ranking but move the 9 so as to preserve the number S at 26. The 1 passes over three members to go to the beginning and hence adds 3 to the score. The 9 must, therefore, proceed to the left over three numbers so as to subtract 3 from the score and we reach

$$1\quad 5\quad 2\quad 3\quad 9\quad 6\quad 7\quad 8\quad 4.$$

Now operate similarly with 2 and 9, reaching

$$1\quad 2\quad 5\quad 9\quad 3\quad 6\quad 7\quad 8\quad 4.$$

Had our 9 been contiguous to the 1 and incapable of moving farther to the left we should have moved the 8 and so on. Proceeding with the process by moving back the 3 and the 9 and 8 we reach

$$1\quad 2\quad 3\quad 9\quad 5\quad 6\quad 8\quad 7\quad 4,$$

and again

$$1\quad 2\quad 3\quad 4\quad 9\quad 8\quad 7\quad 6\quad 5.$$

All the lower numbers 1 to 4 are in the right order and the remainder are in the inverse order. We call this ranking the 'canonical' order for given S (or t). It is not always possible to reduce a given ranking to canonical order, but there cannot be more than one individual out of place.

2. Consider the effect of a series of transformations leading to the canonical form. The first process, that of moving 1 and 9, will increase the value of S for some samples involving 1 but not 9 (leaving the others unchanged), will decrease the value of S for some samples involving 9 but not 1 (leaving the others unchanged), and will, in general, not alter those involving both 1 and 9. Similarly for 2 and 8, and so on. The effect of the transformation is thus to increase the values of S containing the lower numbers 1, 2, 3, etc., and to decrease those containing 9, 8, 7, etc. These values of S are themselves, in the canonical form, the greatest or least as the case may be. Consequently the progress to the canonical form is accompanied by increases in the number of high values of S and increases in the number of lower values, and one might expect the spread of the distribution to tend to a maximum. In the example quoted, the distributions of S in samples of 3 for the successive rankings are:

Values of S	Frequencies f				
0	2	3	3	6	10
1	15	13	16	10	—
2	34	35	29	32	40
3	33	33	36	36	34
Totals	84	84	84	84	84

The sums ΣfS are all equal to 182. The sums ΣfS^2 are respectively 448, 450, 456, 462 and 466, showing the canonical ranking to have the largest variance of the five.

3. There is, however, another way of carrying out this process. If the parent ranking is inverted, τ becomes $-\tau$, but the variance of samples of n drawn from the inverted ranking remains the same, by symmetry. We may then reduce the inverted ranking to its canonical form and reinvert it so that its coefficient is again τ. This ranking we call the inverse canonical form. It will be shown that for large N, when $\tau > 0$ the inverse canonical form yields a larger variance for t than the direct canonical form.

Even in the example already quoted, the inverse canonical ranking (with one member out of place) is

$$3 \quad 4 \quad 2 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9 \quad 1,$$

which has a distribution

Values of S	f
0	2
1	27
2	10
3	45
Total	84

The sum $\Sigma f S^2$ is now 472, which is greater than the previous maximum 466.

4. Consider the canonical case when there are N members altogether, R at the beginning in the right order, and $N-R$ in the inverse order. If we select $n-j$ members from the R and j from the $N-R$ the value of S for the sample of n is $\frac{1}{2}n(n-1)-\frac{1}{2}j(j-1)$, and the relative frequency of $U = \frac{1}{2}n(n-1)-S$ is $\binom{R}{n-j}\binom{N-R}{j}\Big/\binom{N}{n}$. Now suppose that N tends to infinity and R/N to the ratio p. The relative frequency of $U = \frac{1}{2}j(j-1)$ tends in the limit to

$$\binom{n}{j} p^{n-j}q^j,$$

where $q = 1-p$. The mean value of U is then

$$\sum_0^n \tfrac{1}{2}j(j-1)\binom{n}{j} p^{n-j}q^j = \tfrac{1}{2}n(n-1)\,q^2,$$

and since

$$t = 1-\frac{2U}{\frac{1}{2}n(n-1)},$$

we must have

$$q = \{\tfrac{1}{2}(1-\tau)\}^{\frac{1}{2}}. \tag{4·1 A}$$

The variance of U is

$$\operatorname{var} U = n(n-1)\,pq^2\{nq-\tfrac{1}{2}(1-3q)\}, \tag{4·2 A}$$

and so

$$\operatorname{var} t = 16pq^2\{nq-\tfrac{1}{2}(1-3q)\}/n(n-1). \tag{4·3 A}$$

5. If now the inverted parent ranking is reduced to canonical form, giving ratios p', q' corresponding to p and q, we shall have

$$q' = \sqrt{[\tfrac{1}{2}(1+\tau)]} \tag{5·1 A}$$

and

$$\operatorname{var} t' = 16p'q'^2\{nq'-\tfrac{1}{2}(1-3q')\}/n(n-1). \tag{5·2 A}$$

Then since $q^2+q'^2 = 1$,

$$\operatorname{var} t' - \operatorname{var} t = \frac{16(n+2)}{n(n-1)}(q'-q)(1-q)(1-q'). \tag{5·3 A}$$

When τ is positive, $q' > q$ and $\operatorname{var} t'$ exceeds $\operatorname{var} t$.

This result suggests that the maximum variance may be attained by the inverse canonical ranking when $\tau > 0$ and by the direct canonical ranking when $\tau < 0$. With this choice of parent ranking the variance of t for large n is

$$\operatorname{var} t \sim \frac{4\sqrt{2}}{n}(1+|\tau|)^{\frac{3}{2}}\{1-\sqrt{[\tfrac{1}{2}(1+|\tau|)]}\}. \qquad (5\cdot4\,\mathrm{A})$$

It is interesting to compare (5·4 A) with our upper limit of $2(1-\tau^2)/n$. Their ratio is $\{2(1+|\tau|)\}^{\frac{1}{2}}/[1+\{\tfrac{1}{2}(1+|\tau|)\}^{\frac{1}{2}}]$, which varies from $2(\sqrt{2}-1) = 0\cdot83$ when $\tau = 0$ to 1 when $\tau = 1$. Evidently the upper limit to the variance cannot be much improved, since an actual ranking has been found whose variance approximates to it for all values of τ, when n is not too small.

REFERENCES

Daniels, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika*, **33**, 129.

Kendall, M. G. (1943). *The Advanced Theory of Statistics*, **1**. 2nd ed. 1945. London: Charles Griffin and Co.

TESTING FOR NORMALITY

By R. C. GEARY, *Cambridge University Department of Applied Economics*

1. Introduction

The present communication, one of a series, has two main objectives:

(1) To show that probabilities derived from the well-known analyses of variance and other 'small sample' tables, which postulate universal normality, may differ seriously from the true probabilities when the universes are non-normal, even, in some cases, when the degree of non-normality is not considerable.

(2) To determine the most efficient tests of normality from a wide field of alternative symmetrical tests.

It may be useful to summarize very briefly previous work in so far as it is strictly relevant to this study.* The modern theory may be regarded as having been initiated by Karl Pearson who, in 1895, found the first approximation (i.e. to n^{-1}) to the variances and covariance of $\sqrt{b_1}$ and b_2 for samples drawn at random from any universe and, assuming that the $\sqrt{b_1}$ and b_2 were distributed jointly with normal probability, constructed 'probability ellipses' from which the probability of the same values occurring, had the universe, in fact, been normal, could be inferred very approximately. A considerable advance in moment determination was made by C. C. Craig (1928). In 1929, R. A. Fisher, in inventing cumulants, simple functions of the sample moments, and formulating rules for finding their semi-invariants, developed incidentally a technique for expanding to several terms in $1/n$ the moments of $\sqrt{b_1}$ and b_2 when the universe was normal. This paper was followed soon after by another (1930), fundamental for all succeeding work on this subject, in which R. A. Fisher ingeniously applied combinatorial technique to the finding of exact values of the moments of normal $\sqrt{b_1}$ and b_2, and gave *inter alia* the values of the second, fourth and sixth moments of $\sqrt{b_1}$ and of the first three moments of b_2. The fourth semi-invariant, together with many other normal semi-invariants of b_2, was determined by J. Wishart in 1930, and a further advance in R. A. Fisher's technique was made jointly by R. A. Fisher & J. Wishart in 1930. In 1932 Joseph Pepper gave the eighth normal moment of $\sqrt{b_1}$. Using R. A. Fisher's rules C. T. Hsu and D. N. Lawley in 1940 gave the exact values for normal random samples of the fifth and sixth moments of b_2. Using a method due to R. C. Geary (1933) (applying C. C. Craig's ideas (1928) to the normal problem), R. C. Geary & J. P. G. Worlledge have recently (1946) found the seventh moment of b_2.

So much for moment determination. In 1930, E. S. Pearson used appropriate Pearson-type curves, applied to R. A. Fisher's (1929) approximations of the semi-invariants, to find approximate frequency distributions of $\sqrt{b_1}$ and b_2. From the frequency distributions he computed a table of 1 % and 5 % probability points at intervals for n from 50 to 5000 for $\sqrt{b_1}$ and for n from 100 to 5000 for b_2.

Since at the time the prospect seemed remote of determining the frequency of normal b_2 on which reliance could be reposed for samples of moderate sizes, R. C. Geary (1935)† suggested that the ratio, a, of mean deviation to standard deviation computed from the origin

* An excellent account of the development of moment theory up to the year 1930 was given by J. Wishart (1930).

† The author was informed by M. Fréchet that this test was suggested by Bertrand, but has been unable to check the reference.

might be used as a test of normality, and gave the 1 and 5 % probability points for this test at intervals for normal samples of 6–100. E. S. Pearson compared experimentally Geary's test with b_2 and suggested, for samples so large that comparison could safely be made, that b_2 was probably somewhat more sensitive than a, a suggestion which will be examined theoretically in this communication. In 1935 also, R. C. Geary showed that there was a high (negative) correlation for normal samples between $a(1)$ (see 3·1) and b_2 for normal samples, and argued therefrom that the former should be nearly as efficient as b_2. In 1936, R. C. Geary gave a table of 1, 5 and 10 % probability points of $a(1)$ at intervals for samples of 11–1001. In 1938, a brochure by R. C. Geary & E. S. Pearson was published by the Biometrika Office entitled *Tests of Normality*, giving tables and diagrams of probability points of $a(1)\sqrt{b_1}$ and b_2. There is considerable literature dealing with the effect of universal non-normality on the normal tests, mostly by way of particular numerical examples: a selection of papers on this subject is included in the list of references at the end of the paper.

2. Effect of non-normality

(a) The z-test

The effect of universal non-normality will first be considered in relation to the z-test. If $x_1, x_2, \ldots, x_{n'}$ and $y_1, y_2, \ldots, y_{n''}$ are two independent samples drawn at random from the same universe (normal or non-normal) it is easy to show that, if

$$z = \tfrac{1}{2}\log\frac{n''-1}{n'-1}\frac{\sum\limits_{i=1}^{n'}(x_i-\bar{x})^2}{\sum\limits_{i=1}^{n''}(y_i-\bar{y})^2} = \tfrac{1}{2}\log\frac{s'^2}{s''^2}, \tag{2·1}$$

then

$$\sigma_z^2 = \frac{(\beta_2-1)}{4}\left(\frac{1}{n'}+\frac{1}{n''}\right) = M_2, \tag{2·2}$$

when both n' and n'' are so large that terms in n' and n'' of degree less than -1 are regarded as negligible. This is an obvious generalization of the approximate formula given by R. A. Fisher* for normal samples, namely,

$$\sigma_z^2 = \frac{1}{2}\left(\frac{1}{n'}+\frac{1}{n''}\right) = M_2^0. \tag{2·3}$$

It may be useful also to give formulae for the first and second moments from zero for z when the two random samples are drawn not necessarily from the same universes, though both universes have mean zero and the same variance λ_2:

$$\left.\begin{aligned}
2M_1' = {} & -\frac{1}{2\lambda_2^2}\left(\frac{\lambda_4'}{n'}+\frac{2\lambda_2^2}{n'-1}\right)+\frac{1}{2\lambda_2^2}\left(\frac{\lambda_4''}{n''}+\frac{2\lambda_2^2}{n''-1}\right)+\frac{1}{3\lambda_2^3}\left[\left(\frac{\lambda_6'}{n'^2}-\frac{\lambda_6''}{n''^2}\right)+12\lambda_2\left(\frac{\lambda_4'}{n'^2}-\frac{\lambda_4''}{n''^2}\right)\right. \\
& \left.+4\left(\frac{\lambda_3'^2}{n'^2}-\frac{\lambda_3''^2}{n''^2}\right)+8\lambda_2^3\left(\frac{1}{n'^2}-\frac{1}{n''^2}\right)\right]-\frac{3}{4\lambda_2^4}\left[\frac{(\lambda_4'+2\lambda_2^2)^2}{n'^2}-\frac{(\lambda_4''+2\lambda_2^2)^2}{n''^2}\right]+\ldots, \\
4M_2' = {} & \frac{1}{\lambda_2^2}\left[\left(\frac{\lambda_4'}{n'}+\frac{\lambda_4''}{n''}\right)+2\lambda_2^2\left(\frac{1}{n'-1}+\frac{1}{n''-1}\right)\right] \\
& -\frac{1}{\lambda_2^3}\left[\left(\frac{\lambda_6'}{n'^2}+\frac{\lambda_6''}{n''^2}\right)+12\lambda_2\left(\frac{\lambda_4'}{n'^2}+\frac{\lambda_4''}{n''^2}\right)+4\left(\frac{\lambda_3'^2}{n'^2}+\frac{\lambda_3''^2}{n''^2}\right)+8\lambda_2^3\left(\frac{1}{n'^2}+\frac{1}{n''^2}\right)\right] \\
& +\frac{1}{12\lambda_2^4}\left[\frac{33}{n'^2}(\lambda_4'+2\lambda_2^2)^2+\frac{33}{n''^2}(\lambda_4''+2\lambda_2^2)^2-\frac{6}{n'n''}(\lambda_4'+2\lambda_2^2)(\lambda_4''+2\lambda_2^2)\right]+\ldots,
\end{aligned}\right\} \tag{2·4}$$

* *Statistical Methods for Research Workers*, 8th ed. p. 219.

where the λ's indicate semi-invariants of the two universes of the orders indicated. In these formulae, in effect, terms to order -2 in n', n'' are retained.

When both samples are large the frequency distribution of z will approach normality provided that μ_4 is finite. The effect of universal kurtosis can accordingly be assessed in a very rudimentary manner from (2·2) and (2·3). The z-deviate ζ corresponding to, say, the $2\frac{1}{2}$ % normal probability point is

$$\zeta = 1{\cdot}9600\sqrt{M_2^0}. \tag{2·5}$$

If, however, the universe were not normal and had, *in fact*, a variance M_2 with $\beta_2 \neq 3$, the *actual* probability of a deviation in excess of ζ in absolute value would be, not 0·05, but the normal probability appropriate to a unit variance deviate of $\zeta M_2^{-\frac{1}{2}}$. On this consideration the actual probabilities for different values of β_2, where the assumed probability is 0·05, are shown in the fifth column of Table 1.

Table 1. *Effect on probability of z of change in universal kurtosis, for large samples*

β_2	M_2^0/M_2	$\sqrt{(M_2^0/M_2)}$	$1{\cdot}9600\sqrt{(M_2^0/M_2)}$	Actual probability
1·5	4	2	3·9200	0·000089
2	2	1·4142	2·7718	0·0056
2·5	1·3333	1·1547	2·2632	0·024
3	**1**	**1**	**1·9600**	**0·050**
3·5	0·8000	0·8944	1·7530	0·080
4	0·6667	0·8165	1·6003	0·110
4·5	0·5714	0·7559	1·4816	0·138
5	0·5000	0·7071	1·3859	0·166
5·5	0·4444	0·6667	1·3065	0·191
6	0·4000	0·6325	1·2397	0·215

The table shows that, if the universe from which the samples are drawn has $\beta_2 = 6$, the true probability is about 1 in 5 instead of the assumed 1 in 20. It is, of course, true that universes with so large a kurtosis are unusual. This view cannot be held of the range 2·5–4 for β_2 in which the probability, assumed to be 0·05, can be anything, in fact, from 0·024 to 0·110. Accordingly, if universal kurtosis is markedly negative, use of the standard table masks significant differences; if kurtosis is positive the standard table exaggerates these differences. Unless systematic tests have established that kurtosis is negligible the standard table should not be used for testing significant differences in variance.

The foregoing analysis gives a theoretical explanation of the striking experimental results of E. S. Pearson (1931*b*) working, however, with a test function

$$x = \sum_{i=1}^{n'} (x_i - \bar{x})^2 \Bigg/ \left\{ \sum_{i=1}^{n'} (x_i - \bar{x})^2 + \sum_{i=1}^{n''} (y_i - \bar{y})^2 \right\}$$

and with sample sizes $n' = 5$ and $n'' = 20$, smaller than those contemplated in the present analysis. With 500 samples Pearson showed that when the frequency at the two tails together expected from normal theory was 15·4 (=probability 0·0308) the frequencies actually found in symmetrical universes with $\beta_2 = 2{\cdot}5$, 4·1 and 7·1 respectively were 7, 39 and 47, equivalent to probabilities of 0·014, 0·078 and 0·094.

If tests of normality indicate universal kurtosis, either of two courses might be adopted:

(i) Assume that z is normally distributed with variance M_2 computed from (2·2) with (β_2-3) estimated as k_4/k_2^2 from the sample, k_2 and k_4 being R. A. Fisher's (1929) cumulant functions.

(ii) Enter the standard table, not with z computed from the samples but with $z\sqrt{(M_2^0/M_2)}$, estimating M_2 as in (i).

Both of these procedures are, of course, open to the objection that, unless the samples are extremely large the estimate of β_2 is unlikely to be accurate; the real β_2 might be larger or smaller than the estimate. Any probabilistic inferences should accordingly be accepted with reserve.

It is fortunate that the condition specified in the foregoing paragraphs, namely, that the numbers in the two samples are both large, rarely applies in practical applications. It more usually happens that the number of classes is small, whereas the number per class is relatively large. In this case E. S. Pearson (1931 *b*) has shown the first approximation to σ_z^2 is independent of β_2, from which he inferred that the actual probability when the total number of samples was large was inconsiderably influenced by kurtosis. In view of the foregoing analysis it seemed to the writer desirable to carry the inquiry a stage further.

Suppose, then, that k samples are drawn at random from the same universe, n_j in the jth sample, the total $\sum\limits_j n_j = n$. It is assumed that n is so large that terms in n^{-2} are negligible, that the number of samples k is small, and that all the n_j are of the same order of magnitude as n, i.e. that if

$$n_j = \pi_j n, \quad \sum_{j=1}^{k} \pi_j = 1, \tag{2·6}$$

none of the π_j is negligibly small.

Using R. A. Fisher's cumulant notation with subscript to indicate the sample from which the cumulants were computed, the mean for the jth sample is written k_{1j} and its variance k_{2j}. Then

$$z = \tfrac{1}{2}\log\frac{X}{Y}, \tag{2·7}$$

where
$$(k-1)X = \sum_j n_j(k_{1j}-k_1)^2 = \sum n_j k_{1j}^2 - \frac{1}{n}\Sigma^2 n_j k_{1j},$$

so that
$$\frac{k-1}{n}X = \sum \pi_j(1-\pi_j)\,k_{1j}^2 - 2\sum_{j<j'}\sum \pi_j\pi_{j'}k_{1j}k_{1j'},$$

and
$$(n-k)\,Y = \sum_j (n_j-1)\,k_{2j},$$

so that
$$Y = \Sigma\phi_j k_{2j},$$

where
$$\phi_j = \frac{n_j-1}{n-k}.$$

Without loss of generality let the universal mean be zero and the variance unity. It may easily be shown that

$$EX = EY = 1.$$

Set
$$w = \frac{X}{Y} = \frac{\bar{X}+(X-\bar{X})}{\bar{Y}+(Y-\bar{Y})} = \{1+(X-1)\}\{1+(Y-1)\}^{-1}.$$

Then
$$\left.\begin{aligned} w &= \{1+(X-1)\}\{1-(Y-1)+(Y-1)^2-(Y-1)^3+\ldots\},\\ w^2 &= \{1+(X-1)\}^2\{1-2(Y-1)+3(Y-1)^2-4(Y-1)^3+\ldots\}.\end{aligned}\right\} \tag{2·8}$$

We shall compute the approximate values of Ew and Ew^2, i.e. the values to order n^{-1}; the symbol $\simeq$ denotes 'equal to, to approximation required'. From values of the variances and covariances given by E. S. Pearson (1931 *b*) in his equations (9)–(11), we have

$$\left.\begin{aligned} E(X-1)^2 &= \frac{2}{k-1} + (1-2k+\alpha_{-1}-1)\frac{\lambda_4}{n(k-1)^2}, \\ E(X-1)(Y-1) &\simeq \frac{\lambda_4}{n}, \\ E(Y-1)^2 &\simeq \frac{\lambda_4+2}{n}, \end{aligned}\right\} \tag{2·9}$$

with
$$\alpha_c = \sum_j \pi_j^c.$$

We require
$$\left(\frac{k-1}{n}\right)^2 X^2 \simeq \sum_j \pi_j^2(1-\pi_j)^2 k_{1j}^4 - 4\sum_j\sum_{j'} \pi_j^2(1-\pi_j)\pi_{j'} k_{1j}^3 k_{1j'}$$
$$+2\sum_j\sum_{j'}\pi_j\pi_{j'}(1-\pi_j-\pi_{j'}+3\pi_j\pi_{j'})k_{1j}^2k_{2j'}^2 - 4\sum_j\sum_{j'}\sum_{j''}\pi_j\pi_{j'}\pi_{j''}(1-3\pi_j)k_{1j}^2k_{1j'}k_{1j''},$$
$$Y-1 = \Sigma\phi_j(k_{2j}-1) = \Sigma\phi_j k'_{2j}, \quad \text{say,}$$

remembering that, by definition of cumulants,
$$Ek_{2j} = \lambda_2 = 1.$$

Also
$$(Y-1)^2 = \Sigma\phi_j^2 k_{2j}'^2 + 2\sum_{j>j'}\sum \phi_j\phi_{j'}k'_{2j}k'_{2j'}.$$

It will be useful for what follows to note that
$$\phi_j \simeq \pi_j.$$

Using R. A. Fisher's formulae (1929) for formation of joint semi-invariants of k_1 and k_2, and noting that the k samples are independent, we find from the foregoing

$$\left.\begin{aligned} n(k-1)\,EX(Y-1)^2 &\simeq (k-1)(\lambda_4+2), \\ n(k-1)^2\,EX^2(Y-1) &\simeq 2(k^2-1)\lambda_4, \\ n(k-1)^2\,EX^2(Y-1)^2 &\simeq (k^2-1)(\lambda_4+2). \end{aligned}\right\} \tag{2·10}$$

Then, from (2·8), (2·9), (2·10),

$$\left.\begin{aligned} Ew &\simeq 1+\frac{2}{n}, \\ Ew^2 &\simeq \frac{k+1}{k-1} + \frac{1}{n(k-1)^2}\{6(k^2-1)-(k^2+2k-2-\alpha_{-1})\lambda_4\}. \end{aligned}\right\} \tag{2·11}$$

These are the formulae required. It will be noted

(i) that the terms free of n^{-1} are independent of λ_4, which is equivalent to E. S. Pearson's result (1931 *b*);

(ii) that the formulae (2·11) agree with the normal values

$$\left.\begin{aligned} E_0 w &= \left(1-\frac{2}{n-k}\right)^{-1} \simeq 1+\frac{2}{n}, \\ E_0 w^2 &= \frac{k+1}{k-1}\left(1-\frac{2}{n-k}\right)^{-1}\left(1-\frac{4}{n-k}\right)^{-1} \simeq \frac{k+1}{k-1}\left(1+\frac{6}{n}\right), \end{aligned}\right\} \tag{2·12}$$

to n^{-1} when $\lambda_4 = 0$;

(iii) the approximations at (2·11) are free of λ_3.

The approximations at (2·11) tend to confirm E. S. Pearson's result that, when n is large compared with k, the effect of universal kurtosis is unimportant. It would be useful, however, to compute the approximate true probability for different values of k, n, λ_4 and α_{-1}. For this and for subsequent work the following lemma* will be found useful:

If $f(x)$ and $\phi(x)$ are two frequency densities with semi-invariants L_m and L'_m $(m = 1, 2, \ldots)$, respectively, then, formally,

$$f(x) = \exp\left\{\sum_{m=1}^{\infty} \frac{(L_m - L'_m)}{m!}\left(-\frac{d}{dx}\right)^m\right\}\phi(x). \tag{2·13}$$

For the present application take as generating function ϕ the frequency distribution of w in the normal case, i.e.

$$\phi(w) = \frac{\left(\frac{n-3}{2}\right)!\left(\frac{k-1}{n-k}\right)^{\frac{1}{2}(k-1)}}{\left(\frac{k-3}{2}\right)!\left(\frac{n-k-2}{2}\right)!}\, w^{\frac{1}{2}(k-3)}\left\{1 + \frac{(k-1)\,w}{(n-k)}\right\}^{-\frac{1}{2}(n-1)}, \tag{2·14}$$

and, from (2·11),

$$L_2 - L'_2 \simeq -\frac{(k^2 + 2k - 2 - \alpha_{-1})\,\lambda_4}{n(k-1)^2}. \tag{2·15}$$

Assume that

$$L_m - L'_m \simeq 0 \quad (m \neq 2).$$

Then if the 'normal theory' probability corresponding to the sample value w be p, the approximate 'true' probability, subject to (2·15), will be about $(p + p')$, where p' is given by

$$p' = \frac{(L_2 - L'_2)}{2}\int_w^{\infty} \phi''(w)\,dw = -\frac{(L_2 - L'_2)}{2}\phi'(w). \tag{2·16}$$

The term p', of course, merely corrects for the non-normal term in n^{-1} in the variance of z; it takes no account of corrections due to terms of higher (negative) orders in n or even of non-normal terms in n^{-1} in semi-invariants L_m $(m > 2)$. The calculation is designed merely to show whether the standard table probability requires correction for universal kurtosis; this will appear if p' is of the order of magnitude of p.

(b) *The t-test*

In Geary's 1936 paper the expansion to terms in n^{-2} of the first *four* moments of t, where

$$t = n^{\frac{1}{2}}k_1/k_2^{\frac{1}{2}}, \tag{2·17}$$

were given. Following are the first *six* semi-invariants L of t to the same approximation as in the earlier paper:

$$\left.\begin{aligned}
L_1 &\simeq -\frac{1}{n^{\frac{1}{2}}}\left\{\frac{\lambda_3}{2} + \frac{3}{16n}(2\lambda_3 - 2\lambda_5 + 5\lambda_3\lambda_4)\right\} + \ldots,\\
L_2 &\simeq 1 + \tfrac{1}{4}(8 + 7\lambda_3^2)\,n^{-1} + (6 - 2\lambda_4 - \tfrac{3}{8}\lambda_3^2 - \tfrac{45}{8}\lambda_3\lambda_5 + \tfrac{177}{16}\lambda_3^2\lambda_4)\,n^{-2},\\
L_3 &\simeq -2\lambda_3 n^{-\frac{1}{2}} - (9\lambda_3 - 3\lambda_5 + \tfrac{15}{4}\lambda_3\lambda_4 + \tfrac{83}{8}\lambda_3^3)\,n^{-\frac{3}{2}},\\
{}^{\dagger}L_4 &\simeq (6 - 2\lambda_4 + 12\lambda_3^2)\,n^{-1} + (54 - 18\lambda_4 + 4\lambda_6 + 75\lambda_3^2 - 63\lambda_3\lambda_5 - 6\lambda_4^2 + 81\lambda_3^2\lambda_4 + \tfrac{699}{8}\lambda_3^4)\,n^{-2},\\
L_5 &\simeq -(60\lambda_3 - 6\lambda_5 - 20\lambda_3\lambda_4 + 105\lambda_3^3)\,n^{-\frac{3}{2}},\\
L_6 &\simeq (240 - 120\lambda_4 + 577\tfrac{1}{2}\lambda_3^2 + 16\lambda_6 - 210\lambda_3\lambda_5 - 150\lambda_3^2\lambda_4 + 1200\lambda_3^4)\,n^{-2}.
\end{aligned}\right\} \tag{2·18}$$

Throughout this subsection we take $\lambda_m = \lambda'_m/\lambda_2'^{\frac{1}{2}m}$,

* Due to Charlier and termed the "Differential Series" by the Scandinavian School.

† 1936 formula corrected.

where the λ'_m are the semi-invariants of the parent universe. For these expressions terms in $n^{-\frac{3}{2}}$ are neglected. They were derived from the moments (from zero) M'_i of t, which were obtained by the method described in the 1936 paper. It will be noted that, to the approximation used, the expressions involve only the first six semi-invariants of the parent universe. When the parent universe is normal all the λ_i $(i > 2)$ are zero. The magnitude of the numerical coefficients in the foregoing approximate expressions for the L_i indicate that, when the universal values of the λ_i, particularly those of uneven order, are not very small, the frequency distribution of t may differ appreciably from the classical Gosset-Fisher (1908, 1925) distribution.

The formal Gram-Charlier expression for the frequency of t could, of course, be written down at once from (2·18). It is doubtful, however, if the Gaussian can be regarded as the most appropriate generating function for the frequency of t because, even when the parent universe is normal, the semi-invariants T'_{2m} of the higher even orders are large for moderate values of n. For example,

$$L'_4/L'^2_2 = 6/(n-5), \quad L'_6/L'^3_2 = 240/(n-5)(n-7).$$

It is proposed to use (2·13) for finding the approximate frequency with

$$\phi(t) = T(t;n) = \left(\frac{n-2}{2}\right)!\left(1+\frac{t^2}{n-1}\right)^{-\frac{1}{2}n} \bigg/ \left(\frac{n-3}{2}\right)!\,(\pi\,\overline{n-1})^{\frac{1}{2}}, \tag{2·19}$$

the Gosset-Fisher frequency. Let

$$T_1(t;n) = \left(1+\frac{t^2}{n-1}\right)^{-\frac{1}{2}n}. \tag{2·20}$$

It can easily be shown that the rth derivative (in t) of T_1 is

$$T_1^{(r)}(t;n) = (-)^r \frac{(n+r-1)!}{(n-1)!\,(n-1)^r}\left\{t^r - n_1\frac{r(r-1)}{2}t^{r-2} + n_2\frac{r(r-1)(r-2)(r-3)}{2.4}t^{r-4} - n_3\frac{r(r-1)(r-2)(r-3)(r-4)(r-5)}{2.4.6}t^{r-6}+\ldots\right\}\left(1+\frac{t^2}{n-1}\right)^{-\frac{1}{2}(n+2r)}, \tag{2·21}$$

with $$n_1 = \frac{n-1}{n+1}, \quad n_2 = \frac{(n-1)^2}{(n+1)(n+3)}, \quad n_3 = \frac{(n-1)^3}{(n+1)(n+3)(n+5)}, \quad \text{etc.}$$

Note that (2·21) assumes the Hermite form when $n = \infty$.

The theory will now be applied to particular examples using in all cases $n = 10$. The universes will be assumed to belong to the Karl Pearson system, so that (M. G. Kendall, 1941) the values of λ_5 and λ_6 can be derived (given λ_3 and λ_4) from the following equations:

$$\left.\begin{aligned}(1+4\eta)\lambda_3 + 2\xi &= 0,\\ (1+5\eta)\lambda_4 + 3\xi\lambda_3 + 6\eta &= 0,\\ (1+6\eta)\lambda_5 + 4\xi\lambda_4 + 24\eta\lambda_3 &= 0,\\ (1+7\eta)\lambda_6 + 5\xi\lambda_5 + 10\eta(4\lambda_4 + 3\lambda_3^2) &= 0.\end{aligned}\right\} \tag{2·22}$$

From the first two equations

$$\eta = (2\lambda_4 - 3\lambda_3^2)/(-10\lambda_4 + 12\lambda_3^2 - 12),$$

which, substituted in the first equation of (2·22), gives ξ. The values of ξ and η, substituted in the third and fourth equations, give λ_5 and λ_6. From (2·18), the L'_i being the semi-invariants

when the parent universe is normal (i.e. the values found when all the λ's are set equal to zero),

$$\left.\begin{aligned} L_1 - L_1' &\simeq J_1 n^{-\frac{1}{2}} + K_1 n^{-\frac{3}{2}}, & L_4 - L_4' &\simeq J_4 n^{-1} + K_4 n^{-2}, \\ L_2 - L_2' &\simeq J_2 n^{-1} + K_2 n^{-2}, & L_5 - L_5' &\simeq \qquad\qquad K_5 n^{-\frac{3}{2}}, \\ L_3 - L_3' &\simeq J_3 n^{-\frac{1}{2}} + K_3 n^{-\frac{3}{2}}, & L_6 - L_6' &\simeq \qquad\qquad K_6 n^{-2}. \end{aligned}\right\} \tag{2·23}$$

The J and K are the terms in the λ in (2·18). To n^{-2} (i.e. ignoring $n^{-\frac{5}{2}}$) the frequency generated from T of (2·19) is as follows:

$$\begin{aligned} f(t) = T &+ n^{-\frac{1}{2}}\left\{J_1 D + \frac{J_3}{6} D^3\right\} + n^{-1}\left\{\frac{D^2}{2}(J_2 + J_1^2) + \frac{D^4}{24}(J_4 + 4J_1 J_3) + \frac{D^6}{72} J_3^2\right\} \\ &+ n^{-\frac{3}{2}}\left\{K_1 D + \frac{D^3}{6}(K_3 + 3J_1 J_2 + J_1^3) + \frac{D^5}{120}(K_5 + 5J_1 J_4 + 10J_2 J_3 + 10J_1^2 J_3)\right. \\ &\left. + \frac{D^7}{144}(J_3 J_4 + 2J_1 J_3^2) + \frac{J_3^3}{1296} D^9\right\} + n^{-2}\left\{\frac{D^2}{2}(K_2 + 2J_1 K_1) + \frac{D^4}{24}(K_4 + 4J_1 K_3\right. \\ &+ 4J_3 K_1 + 3J_2^2 + 6J_1^2 J_2 + J_1^4) + \frac{D^6}{720}(K_6 + 6J_1 K_5 + 20J_3 K_3 + 15J_2 J_4 \\ &+ 60J_1 J_2 J_3 + 15J_1^2 J_4 + 20J_1^3 J_3) + \frac{D^8}{11{,}520}(16J_3 K_5 + 10J_4^2 + 80J_2 J_3^2 \\ &\left. + 80J_1 J_3 J_4 + 80J_1^2 J_3^2) + \frac{D^{10}}{5184}(3J_3^2 J_4 + 4J_1 J_3^3) + \frac{J_3^4}{31{,}104} D^{12}\right\}, \end{aligned} \tag{2·24}$$

with
$$D^h = \left(-\frac{d}{dt}\right)^h T.$$

To n^{-1}, (2·24) agrees with the formula given by M. S. Bartlett (1935), in which, however, there is a small and obvious slip in a sign. The law of formation of the numerical coefficients of (2·24) is evident; for instance, the numerical coefficient of $D^8 J_2 J_3^2$ is $1/144 = 1/2!\,3!^2 2!$. The integrals $\int_t^\infty$ and $\int_{-\infty}^{-t}$ $(t > 0)$ are found by reducing the exponent of D by unity, as follows:

$$\int_{-\infty}^{-t} D\,dt = -T, \quad \int_t^\infty D^{2m}\,dt = \int_{-\infty}^{-t} D^{2m}\,dt = D^{2m-1}, \quad \int_t^\infty D^{2m+1}\,dt = -\int_{-\infty}^{-t} D^{2m+1}\,dt = D^{2m}. \tag{2·25}$$

In normal theory the upper and lower $2\frac{1}{2}\%$ points of t are $\pm 2{\cdot}262$ for $n = 10$. Table 2 shows the 'true' probabilities, i.e. the value of

$$\int_{-\infty}^{-2\cdot262} f(t)\,dt \tag{2·26}$$

for parent universes specified by λ_3, λ_4, using (2·24).

There are two observations to be made on the results presented in this table. The first is that, despite the considerable number of terms (shown at (2·24)) included in the probability expansion, the values found in the successive terms cannot be regarded as satisfactorily convergent for so small a sample as 10, and, of course, the convergence disimproves with increasing $\sqrt{\beta_1}$. Taken all together, however, they seem consistent and significant. The second observation is that attention was confined to the negative 'tail' of the distribution. It may be assumed that, in all cases, the distortion would be very considerably less marked if regard were had to the probability for $|t| > 2{\cdot}262$. Actually for universe 3 the probability

is 0·056, not significantly different from the normal theory probability of 0·05. In justification of the attitude adopted above, the point might be put as follows:

We decide to accept the hypothesis that the universal mean is zero provided that the value of t found from the particular sample satisfies $t_0 \leqslant t \leqslant t_1$, where

$$\text{Prob}\,(t < t_0) = \text{Prob}\,(t > t_1) = 0{\cdot}025.$$

The table is designed to show that if the parent universe is markedly asymmetrical the range (t_0, t_1) may differ appreciably from $-t_0 = t_1 = 2{\cdot}262$.

Table 2. *Probabilities of t less than* $-2{\cdot}262$ *for samples of* 10 *for seven universes*

Universe	$\lambda_3 = \sqrt{\beta_1}$	$\lambda_4 = \beta_2 - 3$	Probability
Normal	0	0	0·025
2	0	1	0·024
3	1/2	0	0·041
4	$1/\sqrt{2}$	1/2	0·047
5	1	0	0·072?
6	1	1	0·086?
7	1/2	1/2	0·043

As anticipated by earlier work (W. S. Gosset, 1908; R. C. Geary, 1936), the table shows that the distortion is slight for symmetrical universes; even when $\lambda_4 = 1$ (and $\lambda_3 = 0$) the probability (0·024) is practically identical with the normal value. There can be little doubt that the standard table probabilities can be seriously at variance with the true probabilities when the universes from which the samples are drawn are markedly asymmetrical.

(c) *Difference of means*

R. A. Fisher's (1925) test of significance

$$t = \frac{(k_1' - k_1'')\sqrt{(n' + n'' - 2)}}{\{(n'-1)k_2' + (n''-1)k_2''\}^{\frac{1}{2}}} \sqrt{\frac{n'n''}{(n'+n'')}}, \tag{2·27}$$

for the difference of averages k_1' and k_1'' in normal theory for random samples numbering n' and n'' is, of course, a particular case of the analysis of variance considered in § (*a*) above. The second cumulants are k_2' and k_2''. It is assumed that the unknown universal means and variances are equal. Suppose now that the random samples in reality have been derived from universes in which the means are equal but the other semi-invariants λ_i' and λ_i'' are not necessarily zero for $i \geqslant 2$, or even necessarily equal. Since the universal means are assumed equal, without loss of generality we may take $\lambda_1' = \lambda_1'' = 0$. This general mathematical model seems to be the correct one; we are not trying to determine the probability of the samples being derived from the *same universe* but rather if they could conceivably have been drawn from universes with the *same arithmetic mean*, however much they may differ otherwise. The correctness or otherwise of the concept may be considered in relation to, say, the problem of deciding from two random samples which of two types of fertilizer is to be preferred from yield observations on a given crop on a given kind of land. Undoubtedly the prime problem will be that of ascertaining which is probably the better yielding (i.e. whether the arithmetic means are significantly different). Of considerably less importance is the

question of which fertilizer is the more variable; of less importance still is the question of deciding, say, whether with approximately equal yields one universe is symmetrical and the other markedly asymmetrical. The point is that the question of the equality of universal means should be considered without assuming that the other semi-invariants in the universes from which the samples have been drawn are necessarily equal. This essentially is also the viewpoint in R. A. Fisher's randomization method.

Expanding the denominator of (2·27) in terms of $(k_2' - \lambda_2')$ and $(k_2'' - \lambda_2'')$ and computing therefrom the first few terms of the first four moments of t, we find the following approximations to the first four semi-invariants:

$$\left.\begin{aligned}
AL_1 &\simeq -\frac{(\lambda_3' - \lambda_3'')}{2(n'\lambda_2' + n''\lambda_2'')},\\
A^2L_2 &\simeq \left(\frac{\lambda_2'}{n'} + \frac{\lambda_2''}{n''}\right)\left(1 + \frac{2n'\lambda_2'^2 + n''\lambda_2''^2}{n'\lambda_2' + n''\lambda_2''}\right)\\
&\quad + \frac{(n'^2 - n''^2)(\lambda_4'\lambda_2'' - \lambda_4''\lambda_2')}{n'n''(n'\lambda_2' + n''\lambda_2'')^2} + \frac{7(\lambda_3' - \lambda_3'')^2}{4(n'\lambda_2' + n''\lambda_2'')^2},\\
A^3L_3 &\simeq \frac{\lambda_3'}{n'^2} - \frac{\lambda_3''}{n''^2} - \frac{3(\lambda_3' - \lambda_3'')}{(n'\lambda_2' + n''\lambda_2'')}\left(\frac{\lambda_2'}{n'} + \frac{\lambda_2''}{n''}\right),\\
A^4L_4 &\simeq \frac{6(n'\lambda_2'^2 + n''\lambda_2''^2)}{(n'\lambda_2' + n''\lambda_2'')^2}\left(\frac{\lambda_2'}{n'} + \frac{\lambda_2''}{n''}\right)^2 - 6\left(\frac{\lambda_3'}{n'^2} - \frac{\lambda_3''}{n''^2}\right)\frac{(\lambda_3' - \lambda_3'')}{(n'\lambda_2' + n''\lambda_2'')}\\
&\quad + \frac{18(\lambda_3' - \lambda_3'')^2}{(n'\lambda_2' + n''\lambda_2'')^2}\left(\frac{\lambda_2'}{n'} + \frac{\lambda_2''}{n''}\right) + \frac{\lambda_4'}{n'^3} + \frac{\lambda_4''}{n''^3}\\
&\quad - 3\left(\frac{\lambda_2'}{n'} + \frac{\lambda_2''}{n''}\right)\frac{\{\lambda_4'(n'n''\lambda_2' + 2n''^2 - n'^2\lambda_2'') + \lambda_4''(n'n''\lambda_2'' + 2n'^2 - n''^2\lambda_2')\}}{n'n''(n'\lambda_2' + n''\lambda_2'')^2},
\end{aligned}\right\} \qquad (2\cdot28)$$

with
$$A = \left\{\left(\frac{n' + n''}{n'n''}\right)\frac{(\lambda_2' n' - 1 + \lambda_2'' n'' - 1)}{(n' + n'' - 2)}\right\}^{\frac{1}{2}}.$$

Using formula (2·24) to the term in n^{-1} with the Gosset-Fisher function again as generating function, Table 3 shows rough approximations, for four examples, to the 'true' probability of values of $t \leqslant \tau$, where τ is the (negative) value for probability 0·025 from the normal table, and $\lambda_2' = \lambda_2'' = 1$. When the two samples are drawn from different universes the distortion can accordingly be considerable. The third example suggests that if the universes are the same the distortion is small, a result to be anticipated from the fact (apparent from (2·28)) that, to the approximation used, the first two semi-invariants are equal to their normal theory values; this theory confirms the experimental results of E. S. Pearson & N. K. Adyanthaya (1929).

Table 3

Example	n'	n''	λ_3'	λ_3''	λ_4'	λ_4''	Probability
1	12	4	1	−1	1	−1	0·045
2	18	6	1	−1	1	−1	0·041
3	7	4	$1/\sqrt{2}$	$1/\sqrt{2}$	1/2	1/2	0·027
4	10	6	1	0	1	0	0·036

It should be remarked that the probabilities in Table 3 (as well as in Table 2) are merely rough approximations—the samples used are far too small for the results to have any pretension to accuracy. The object has been merely to show that the actual probability *could* be considerably at variance with that shown in the standard table, for small samples.

3. SUFFICIENT CONDITIONS FOR APPROACH TO NORMALITY OF $a(c)$ WITH INCREASING n

The remainder of the paper deals with the field of symmetrical tests of normality, homogeneous of degree zero, represented by (3·1). It is essential to establish the conditions of approach to normality of the frequency distribution of $a(c)$ as the sample number increases.

Let
$$a(c) = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|^c \Big/ \left\{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}^{\frac{1}{2}c}, \tag{3·1}$$

where $\bar{x} = \Sigma x_i/n$ and c is non-negative. It will be shown in succession that, subject to stated conditions, with increasing n,

(i) the frequency distribution of
$$a_1(c) = \frac{1}{n}\Sigma|x_i|^c \Big/ \left\{\frac{1}{n}\Sigma x_i^2\right\}^{\frac{1}{2}c} \tag{3·2}$$
tends towards normality, and

(ii) the frequency distribution of $a_1(c)$ tends towards that of $a(c)$ and hence towards normality.

It is assumed, without loss of generality, that the universal mean of the universe from which the sample of n is drawn is zero. Denote the kth absolute moment from zero by $\mu_{|k|}$, k not being necessarily an integer. Given a positive quantity ϵ arbitrarily small, $\omega(\epsilon)$ can be found so that
$$\text{Prob}\left\{\left|\frac{1}{n}\Sigma(|x_i|^c - \mu_{|c|})\right| < \omega\sqrt{\left(\frac{\mu_{|2c|}-\mu_{|c|}^2}{n}\right)}\right\} > 1-\epsilon, \tag{3·3}$$
$$\text{Prob}\left\{\left|\frac{1}{n}\Sigma(x_i^2 - \mu_2)\right| < \omega\sqrt{\left(\frac{\mu_4-\mu_2^2}{n}\right)}\right\} > 1-\epsilon, \tag{3·4}$$

provided, of course, that $\mu_{|2c|}$ and μ_4 exist. As n increases ω may be envisaged as approaching the normal probability point appropriate to the probability ϵ, since, in the conditions stated, $\Sigma|x_i|^c/n$ and $\Sigma x_i^2/n$ are normally distributed in the limit. For samples which satisfy the inequality in the brackets { } at (3·4) and if n is so large that
$$\omega\sqrt{\left(\frac{\mu_4}{n}\right)} < \mu_2,$$

the denominator of (3·2) can be expanded to three terms (including the remainder) by Taylor's theorem, so that $a_1(c)$ may be written
$$a_1(c) = \mu_{|c|}\mu_2^{-\frac{1}{2}c}\left\{1+\frac{1}{n}\Sigma\left(y_i - \frac{c}{2}z_i\right) - \frac{c}{2n^2}\Sigma y_i\,\Sigma z_i + \frac{c(c+2)}{8n^2}(\Sigma z_i)^2\left(1+\frac{1}{n}\Sigma y_i\right)X\right\}, \tag{3·5}$$

with
$$y_i = (|x_i|^c - \mu_{|c|})/\mu_{|c|},$$
$$z_i = (x_i^2 - \mu_2)/\mu_2,$$
$$X = \mu_2^{\frac{1}{2}c+2}\left\{\mu_2 + \frac{\theta}{n}\Sigma(x_i^2-\mu_2)\right\}^{-\frac{1}{2}(c+4)} \quad (0<\theta<1).$$

With probability exceeding $(1-\epsilon)$ it is evident, from (3·4), that X is maximized by

$$\left(1-\frac{\omega}{\mu_2}\sqrt{\frac{\mu_4}{n}}\right)^{-\frac{1}{2}(c+4)}.$$

It will suffice, for the present purpose, to infer that

$$|X|<\kappa,$$

where κ is a constant independent of n. We have now

$$E\frac{1}{n}\Sigma\left(y_i-\frac{c}{2}z_i\right)=0.$$

Set
$$\sigma^2=E\frac{1}{n^2}\left\{\Sigma\left(y_i-\frac{c}{2}z_i\right)\right\}^2$$

$$=\frac{1}{n}\left\{\frac{\mu_{|2c|}}{\mu_{|c|}^2}-\frac{c\mu_{|c+2|}}{\mu_{|c|}\mu_2}+\frac{c^2}{4}\frac{\mu_4}{\mu_2^2}-\left(\frac{c}{2}-1\right)^2\right\}, \tag{3·6}$$

and
$$\frac{1}{\sigma}\left(\frac{\mu_2^{\frac{1}{2}c}a_1(c)}{\mu_{|c|}}-1\right)-\frac{1}{n\sigma}\Sigma\left(y_i-\frac{c}{2}z_i\right)=u, \tag{3·7}$$

with
$$u=-\frac{c}{2n^2\sigma}\Sigma y_i\,\Sigma z_i+\frac{c(c+2)}{8n^2\sigma}(\Sigma z_i)^2\left(1+\frac{1}{n}\Sigma y_i\right)X. \tag{3·8}$$

For samples which satisfy the inequalities in { } at (3·3) and (3·4) and hence with a probability exceeding $(1-2\epsilon)$, we have

$$|u|<\frac{c\omega^2}{2\sigma}\frac{\sqrt{(\mu_{|2c|}\mu_4)}}{n\mu_{|c|}\mu_2}+\frac{c(c+2)\kappa\omega^2\mu_4}{8\sigma\quad n\mu_2^2}\left(1+\frac{\omega}{\mu_{|c|}}\sqrt{\frac{\mu_{|2c|}}{n}}\right)<\frac{\xi}{\sqrt{n}}, \tag{3·9}$$

where ξ is independent of n. Or, briefly,

$$\text{Prob}\left\{|u|<\frac{\xi}{\sqrt{n}}\right\}>1-2\epsilon, \tag{3·10}$$

so that u tends in probability towards zero with $1/n$. Now (3·7) may be written in the form $u=Y'-Y$, where Y' and Y are the respective terms on the left side. If A be any number and F the total probability function, a well-known lemma (Fréchet, 1937, p. 164) shows that

$$|F_{Y'}(A)-F_Y(A)|\leqslant\left\{F_Y\left(A+\frac{\xi}{\sqrt{n}}\right)-F_Y\left(A-\frac{\xi}{\sqrt{n}}\right)\right\}+2\epsilon, \tag{3·11}$$

using (3·10). Hence the frequency distribution of

$$Y'=\frac{1}{\sigma}\left(\frac{\mu_2^{\frac{1}{2}c}a_1(c)}{\mu_{|c|}}-1\right) \tag{3·12}$$

tends towards that of
$$Y=\frac{1}{n\sigma}\Sigma\left(y_i-\frac{c}{2}z_i\right) \tag{3·13}$$

at every continuity point of the latter frequency, as n tends towards infinity. But Y, from (3·13), is the simple average of n random measures, and its frequency must tend towards normality provided that its standard deviation exists; from (3·6) it is evident that σ is finite provided that $\mu_{|k|}$, where k is the greater of $2c$ and 4, is finite. Here and in the remainder of this section it will be useful to remember that if $\mu_{|k|}$ exists so does $\mu_{|k'|}$ for $0\leqslant k'\leqslant k$.

To prove that the frequency distribution of $a(c)$ tends towards that of $a_1(c)$ and hence towards normality with increasing n it will be shown that $\Sigma\,|\,x_i-\bar{x}\,|^c/n$ tends in probability towards $\Sigma\,|\,x_i\,|^c/n$. Two cases will be considered separately: (1) $c \geqslant 1$, (2) $1 > c \geqslant 0$.

Case (1). $c \geqslant 1$

For values of x_i for which $|\,x_i\,| \geqslant |\,\bar{x}\,|$,

$$|\,x_i-\bar{x}\,|^c - |\,x_i\,|^c = \pm c\bar{x}\,|\,x_i-\theta\bar{x}\,|^{c-1} \quad (0<\theta<1)$$

and when $|\,x_i\,| < |\,\bar{x}\,|$,

$$\big|\,|\,x_i-\bar{x}\,|^c - |\,x_i\,|^c\,\big| \leqq (2^c+1)\,|\,\bar{x}\,|^c.$$

Hence

$$\frac{1}{n}\left|\sum_{i=1}^{n}(|\,x_i-\bar{x}\,|)^c-(|\,x_i\,|^c)\right| < |\,\bar{x}\,|\left(\frac{B}{n}\sum_{i=1}^{n}|\,x_i\,|^{c-1}+C\,|\,\bar{x}\,|^{c-1}\right), \tag{3·14}$$

B and C being independent of the x_i and n but depending on c. With ϵ arbitrarily small ω can be found so that

$$\left.\begin{aligned}&\text{Prob}\left\{|\,\bar{x}\,| < \omega\sqrt{\frac{\mu_2}{n}}\right\} > 1-\epsilon,\\ &\text{Prob}\left\{\left|\frac{1}{n}\Sigma(|\,x_i\,|^{c-1}-\mu_{|c-1|})\right| < \omega\sqrt{\frac{\mu_{|2c-2|}-\mu^2_{|c-1|}}{n}}\right\} > 1-\epsilon.\end{aligned}\right\} \tag{3·15}$$

Hence, from (3·14) and (3·15), if μ_2 and $\mu_{|2c-2|}$ exist,

$$\text{Prob}\left\{\left|\frac{1}{n}\Sigma\,|\,x_i-\bar{x}\,|^c-\frac{1}{n}\Sigma\,|\,x_i\,|^c\right| < B'\frac{\omega\mu_{|c-1|}\mu_2^{\frac{1}{2}}}{\sqrt{n}}\right\} > 1-2\epsilon$$

for n sufficiently large the constant B' depending on c but not on n. Hence for $c \geqq 1$, $\Sigma\,|\,x_i-\bar{x}\,|^c/n$ tends in probability towards $\Sigma\,|\,x_i\,|^c/n$. Incidentally, this proves that $\{\Sigma(x_i-\bar{x})^2/n\}^{\frac{1}{2}c}$ tends in probability towards $\{\Sigma x_i^2/n\}^{\frac{1}{2}c}$, the latter two expressions representing respectively the denominators of $a(c)$ and $a_1(c)$.

Case (2). $1 > c \geqslant 0$

Let $\bar{x}$ satisfy a probabilistic inequality identical in form with the first equation of (3·15) and let γ be any positive quantity, fixed once for all. Let n (presently to be defined further) be so large that

$$\gamma > \omega\sqrt{\frac{\mu_2}{n}}.$$

Then

$$\frac{1}{n}\sum_{i=1}^{n}(|\,x_i-\bar{x}\,|^c-|\,x_i\,|^c) = \frac{1}{n}\left(\sum_{|x_i|\geqq\gamma}{}' + \sum_{|x_i|<\gamma}{}''\right)(|\,x_i-\bar{x}\,|^c-|\,x_i\,|^c). \tag{3·16}$$

When $|\,x_i\,| \geqslant \gamma$ (i.e. in Σ'),

$$|\,x_i-\bar{x}\,|^c - |\,x_i\,|^c = \pm c\bar{x}\,|\,x_i-\theta\bar{x}\,|^{c-1} \quad (0<\theta<1),$$

so that

$$\text{Prob}\left\{\big|\,|x_i-\bar{x}\,|^c-|\,x_i\,|^c\,\big| < c\omega\left(\gamma-\omega\sqrt{\frac{\mu_2}{n}}\right)^{c-1}\sqrt{\frac{\mu_2}{n}}\right\} > 1-\epsilon. \tag{3·17}$$

When $|\,x_i\,| < \gamma$ (i.e. in Σ''), given η arbitrarily small and positive, n can be found so that

$$\big|\,|\,x_i-\bar{x}\,|^c-|\,x_i\,|^c\,\big| < \eta, \tag{3·18}$$

when

$$|\,\bar{x}\,| < \omega\sqrt{\frac{\mu_2}{n}},$$

since $|\,x\,|^c$ $(c>0)$ is uniformly continuous in Σ''. We then have

$$\text{Prob}\{\big|\,|\,x_i-\bar{x}\,|^c-|\,x_i\,|^c\,\big| < \eta\} > 1-\epsilon. \tag{3·19}$$

Combining (3·17) and (3·19), it may be inferred that

$$\text{Prob}\left\{\left|\frac{1}{n}\Sigma\,|x_i-\bar{x}|^c-\frac{1}{n}\Sigma\,|x_i|^c\right|<c\omega\left(\gamma-\omega\sqrt{\frac{\mu_2}{n}}\right)^{c-1}\sqrt{\frac{\mu_2}{n}}+\eta\right\}>1-2\epsilon, \tag{3·20}$$

the first term of the upper limit in { } tending to zero as n tends towards infinity, and ϵ and η being arbitrarily small

We have accordingly shown that the numerator and denominator of $a(c)$ tends in probability towards those of $a_1(c)$. Hence $a(c)$ tends in probability towards $a_1(c)$. Hence, using the lemma cited at (3·11), the total frequency of $a(c)$ tends towards that of $a_1(c)$ which tends towards normality as n tends towards infinity. Finally:

If $c \geqslant 0$ the frequency distribution of $a(c)$, given by (3·1), *tends towards normality as n tends towards infinity provided that $\mu_{|k|}$, where k is the greater of $2c$ and 4, is finite.*

It seems likely that an analogous theorem can be proved for $0>c>-\frac{1}{2}$; we shall not, however, be concerned in this communication with negative values of c.

4. Moments of $a(c)$ for normal samples

While it will be shown in later sections that, with indefinitely large samples, $\sqrt{b_1}$ and b_2 are the most efficient tests of asymmetry and kurtosis, respectively, it by no means follows that other tests are inefficient or that they may not be useful supplements in cases in which the prime tests are indecisive as to the probable non-normality of a given sample. It is accordingly proposed to give here close approximations to the first four moments (from the origin) of $a(c)$ (given by (3·1)) for normal random samples of n.

For normal samples (R. A. Fisher, 1929; R. C. Geary, 1933)

$$M'_k\{a(c)\} = E\{a(c)\}^k = E\left\{\frac{1}{n}\sum_{i=1}^{n}|x_i-\bar{x}|^c\right\}^k \bigg/ E\left\{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}^{\frac{1}{2}ck}. \tag{4·1}$$

The exact value of the denominator is, of course, known, for

$$E\left\{\frac{1}{n}\Sigma(x_i-\bar{x})^2\right\}^{\frac{1}{2}k'} = \left(\frac{n-1}{n}\right)^{\frac{1}{2}k'} Es^{k'} = \left(\frac{2}{n}\right)^{\frac{1}{2}k'}\left(\frac{n+k'-3}{2}\right)!\bigg/\left(\frac{n-3}{2}\right)!, \tag{4·2}$$

since, as usual, $(n-1)s^2 = \Sigma(x_i-\bar{x})^2$. It will be useful to expand $\log_e Es^{k'}$ with $k' = ck$ using Stirling's formula in (4·2):

$$\begin{aligned}\log_e Es^{k'} &= \frac{k'}{2}\log\frac{2}{n-1}+\log\left(\frac{n+k'-3}{2}\right)!-\log\left(\frac{n-3}{2}\right)!\\ &= \frac{(k'^2-2k')}{4(n-1)}-\frac{k'(k'-1)(k'-2)}{12(n-1)^2}+\frac{k'^2(k'-2)^2}{24(n-1)^3}-\frac{k'(k'-1)(k'-2)(3k'^2-6k'-4)}{120(n-1)^4}\\ &\quad+\frac{k'^2(k'-2)^2(k'^2-2k'-2)}{60(n-1)^5}-\frac{k'(k'-1)(k'-2)(3k'^4-12k'^3+24k'+16)}{252(n-1)^6}\\ &\quad+\frac{k'^2(k'-2)^2(3k'^4-12k'^3-4k'^2+32k'+32)}{336(n-1)^7},\end{aligned} \tag{4·3}$$

which checks for $k' = 1$ to $(n-1)^{-7}$ with Geary (1935, p. 354). Take

$$v(c) = \frac{1}{n}\sum_{i=1}^{n}|z_i|^c, \tag{4·4}$$

with

$$z_i = x_i-\bar{x}.$$

The moments of $v(c)$ will be found exactly as in the case of $c = 1$ (Geary, 1936) from the single or joint normal frequency distributions of $(z_1, z_2, \ldots)$. We find

$$M_1'\{v(c)\} = \frac{1}{\sqrt{\pi}}\left(\frac{2\,\overline{n-1}}{n}\right)^{\frac{1}{2}c}\left(\frac{c-1}{2}\right)!,$$

$$M_2'\{v(c)\} = \frac{1}{\sqrt{\pi}}\frac{(2\,\overline{n-1})^c}{n^{c+1}}\left(\frac{2c-1}{2}\right)! + \frac{2^c}{\pi}n^{-\frac{1}{2}}(n-1)^{-c}(n-2)^{\frac{1}{2}(2c+1)}\left[\left(\frac{c-1}{2}\right)!\right]^2 \tag{4·5}$$

$$\times\left\{1 + \frac{1}{2!}\left(\frac{c+1}{2}\right)^2\left(\frac{2}{n-1}\right)^2 + \frac{1}{4!}\left(\frac{c+1}{2}\right)^2\left(\frac{c+3}{2}\right)^2\left(\frac{2}{n-1}\right)^4 + \frac{1}{6!}\left(\frac{c+1}{2}\right)^2\left(\frac{c+3}{2}\right)^2\left(\frac{c+5}{2}\right)^2\left(\frac{2}{n-1}\right)^6 + \ldots\right\}. \tag{4·6}$$

For the third moment we write

$$M_3'\{v(c)\} = E\{v(c)\}^3 = \frac{n}{n^3}E\,|z_1|^{3c} + \frac{3n(n-1)}{n^3}E\,|z_1|^{2c}\,|z_2|^c + \frac{n(n-1)(n-2)}{n^3}E\,|z_1|^c\,|z_2|^c\,|z_3|^c$$
$$= A_1 + A_2 + A_3, \tag{4·7}$$

denoting the three terms on the right by A_1, A_2, A_3 respectively. Then

$$A_1 = \frac{1}{\sqrt{\pi}}\left(\frac{3c-1}{2}\right)!(2\,\overline{n-1})^{\frac{3}{2}c}\,n^{-\frac{1}{2}(4+3c)},$$

$$A_2 = \frac{3\,.\,2^{\frac{3}{2}c}}{\pi}\left(\frac{2c-1}{2}\right)!\left(\frac{c-1}{2}\right)!(n-2)^{\frac{1}{2}(3c+1)}(n-1)^{-\frac{1}{2}c}\,n^{-\frac{3}{2}}$$
$$\times\left\{1 + \frac{(2c+1)(c+1)}{2!(n-1)^2} + \frac{(2c+3)(2c+1)(c+3)(c+1)}{4!(n-1)^4} + \ldots\right\},$$

$$A_3 = \left(\frac{2^c}{\pi}\right)^{\frac{3}{2}}(n-3)^{\frac{1}{2}(3c+2)}(n-2)^{-\frac{1}{2}(3c+1)}(n-1)\,n^{-\frac{3}{2}}\left[\left(\frac{c-1}{2}\right)!\right]^3\left\{1 + \frac{3}{2}\frac{(c+1)^2}{(n-2)^2}\right.$$
$$-\frac{(c+1)^3}{(n-2)^3} + \frac{(c+1)^2(c+3)(7c+9)}{8(n-2)^4} - \frac{(c+3)^2(c+1)^3}{2(n-2)^5}$$
$$\left.+\frac{(c+3)^2(c+1)^2(61c^2+310c+265)}{240(n-2)^6} + \ldots\right\}.$$

Similarly, for the fourth moment,

$$M_4'\{v(c)\} = E\{v(c)\}^4 = \frac{n}{n^4}E\,|z_1|^{4c} + \frac{4n(n-1)}{n^4}E\,|z_1|^{3c}\,|z_2|^c$$
$$+\frac{3n(n-1)}{n^4}E\,|z_1|^{2c}\,|z_2|^{2c} + \frac{6n(n-1)(n-2)}{n^4}E\,|z_1|^{2c}\,|z_2|^c\,|z_3|^c$$
$$+\frac{n(n-1)(n-2)(n-3)}{n^4}E\,|z_1|^c\,|z_2|^c\,|z_3|^c\,|z_4|^c$$
$$= C_1 + C_2 + C_3 + C_4 + C_5 \tag{4·8}$$

with
$$C_1 = \frac{2^{2c}}{\sqrt{\pi}}(n-1)^{2c}\,n^{-2c-3}\left(\frac{4c-1}{2}\right)!,$$

$$C_2 = \frac{2^{2(c+1)}}{\pi}(n-2)^{\frac{1}{2}(4c+1)}(n-1)^{-2c}\,n^{-\frac{3}{2}}\left(\frac{3c-1}{2}\right)!\left(\frac{c-1}{2}\right)!$$
$$\times\left\{1 + \frac{(3c+1)(c+1)}{2!(n-1)^2} + \frac{(3c+3)(3c+1)(c+3)(c+1)}{4!(n-1)^4} + \ldots\right\},$$

$$C_3 = \frac{3\,.\,2^{2c}}{\pi}(n-2)^{\frac{1}{2}(4c+1)}(n-1)^{-2c}n^{-\frac{1}{2}}\left[\left(\frac{2c-1}{2}\right)!\right]^2\left\{1+\frac{(2c+1)^2}{2!(n-1)^2}+\frac{(2c+3)^2(2c+1)^2}{4!(n-1)^4}+\dots\right\},$$

$$C_4 = \frac{3\,.\,2^{2c+1}}{\pi^{\frac{1}{2}}}(n-3)^{2c+1}(n-2)^{-\frac{1}{2}(4c+1)}(n-1)\,n^{-\frac{1}{2}}\left(\frac{2c-1}{2}\right)!\left[\left(\frac{c-1}{2}\right)\right]!^2$$

$$\times\left\{1+\frac{(c+1)(5c+3)}{2(n-2)^2}-\frac{(c+1)^2(2c+1)}{(n-2)^3}+\frac{(c+1)(57c^3+227c^2+255c+81)}{24(n-2)^4}\right.$$

$$\left.-\frac{(2c+1)(c+1)^2(c+3)(5c+9)}{6(n-2)^5}+\dots\right\},$$

$$C_5 = \frac{2^{2c}}{\pi^2}(n-4)^{\frac{1}{2}(4c+5)}(n-3)^{-2c-1}(n-2)(n-1)\,n^{-\frac{1}{2}}\left[\left(\frac{c-1}{2}\right)!\right]^4$$

$$+\left\{1+\frac{3(c+1)^2}{(n-3)^2}-\frac{4(c+1)^3}{(n-3)^3}+\frac{(c+1)^2(7c^2+21c+15)}{(n-3)^4}-\frac{4(c+3)(c+1)^3(2c+3)}{(n-3)^5}\right.$$

$$\left.+\frac{(c+3)(c+1)^2(122c^3+671c^2+1070c+525)}{15(n-3)^6}-\dots\right\}.$$

Formulae (4·5), (4·6), (4·7) and (4·8) were checked from the corresponding formulae for $c = 1$ given in the author's 1936 paper.

From the following section it will be apparent that for indefinitely large samples the most sensitive test of kurtosis of the field $a(c)$ is found for $c = 4$. At the same time it is shown that there is really not much difference in efficiency for values of c in the range $5 \geqslant c > 2$; moreover, the results in § 6 (in which the efficiency of the tests for $c = 4$ and $c = 1$ are compared from the power function viewpoint) suggest that, for samples of moderate size, the superiority, if any at all, of a test using $a(4) = b_2$ over other tests in the series may be even less marked. The disadvantage of $a(4)$ is that its frequency is not known for samples of all sizes; and if we could estimate, with any degree of confidence, the probability points of $a(c)$ for any value or values of $c > 2$ for medium-size samples we might, for practical purposes, dispense with $a(4)$ altogether, since, while we now know one way of solving the problem of determining the exact, or almost exact, frequency distribution of $a(4)$, it must be admitted that the method is extremely tedious. (From the theoretical point of view, however, the $a(4)$ problem must be solved since it remains a challenge to the mathematical skill of statisticians!) It will accordingly be of interest to study the order of magnitude of the semi-invariants of $a(c)$ for c near 2.

Consider the case, for example, of $c = 2{\cdot}4$, not by any means, it is important to observe, the lowest value which would be used for tabulating. In Table 4 the first three moments are given for $n = 25$. The L's represent, of course, the semi-invariants. The values of the functions for $a_1(c)$ (given by (3·2)) for $n = 24$ (i.e. the appropriate number of degrees of freedom for comparison with $a(c)$) are also given. These show that the moments of $a_1(c)$ are very close to those of $a(c)$, which suggests that, when n is not less than, say, 20, the values of B_1, B_2 and corresponding functions of higher orders, if required, for $a_1(c)$ could be used for the determination of the probability points of $a(c)$. This is important from the computational point of view because the algebraic expressions for the normal moments of $a_1(c)$ are exceedingly simple whereas it must be conceded that (4·8) offers a grim prospect for the computer; furthermore, the principal term C_5 is rather slowly convergent unless $n > 50$ or so,

whereas *exact* values for all values of n can readily be found for the moments of $a_1(c)$ for normal samples.

Table 4. *Normal moments, etc., of* $a(c)$ *and* $a_1(c)$ *for* $c = 2{\cdot}4$

	$a(2{\cdot}4)$	$a_1(2{\cdot}4)$
n	25	24
$M_1' = L_1$	1·166252	1·1662524891
M_2'	1·362004	1·362091186
M_3'	1·592841	1·593151615
$M_2 = L_2$	0·001860	0·001946318
$M_3 = L_3$	0·000063	0·000069583
$\sqrt{B_1} = L_3/L_2^{\frac{3}{2}}$	0·80	0·8104

As with (4·1) for $a(c)$, the moments (from the origin) of any order of $a_1(c)$ is the quotient of the moments of the same order for numerator and denominator, assuming that the universal mean is zero and the variance unity. Since the different members x_i of the sample are independent—the difficulty with $a(c)$ is that the $(x_i - \bar{x})$ are *not* independent—for the moments of the numerator of (3·2) we require only

$$E\,|\,x\,|^{k'} = \frac{2}{\sqrt{(2\pi)}}\int_0^\infty dx\, x^{k'} e^{-\frac{1}{2}x^2} = \left(\frac{k'-1}{2}\right)!\,\frac{2^{\frac{1}{2}k'}}{\sqrt{\pi}}, \tag{4·9}$$

and for the denominator

$$Es^{k'} = E\left(\frac{1}{n}\sum_i x_i^2\right)^{\frac{1}{2}k} = \left(\frac{2}{n}\right)^{\frac{1}{2}k'}\left(\frac{n+k'-2}{2}\right)!\Big/\left(\frac{n-2}{2}\right)!. \tag{4·10}$$

The case of $c = 4$ is particularly simple. The first four semi-invariants are as follows:

$$\left.\begin{aligned} L_1 &= M_1' = \frac{3n}{(n+2)},\\ L_2 &= M_2 = \frac{24n^2(n-1)}{(n+2)^2(n+4)(n+6)},\\ L_3 &= M_3 = \frac{1728(n-1)(n-2)n^3}{(n+2)^3(n+4)(n+6)(n+8)(n+10)},\\ L_4 &= \frac{10{,}368n^4(n-1)(30n^4+168n^3-608n^2-2672n+3712)}{(n+2)^4(n+4)^2(n+6)^2(n+8)(n+10)(n+12)(n+14)}. \end{aligned}\right\} \tag{4·11}$$

Moments, etc., for $a_1(c)$ for normal samples of 24 and 50 are contrasted for $c = 2{\cdot}4$ and $c = 4$ in Table 5. The contrast between the values of $\sqrt{B_1}$ and $(B_2 - 3)$ respectively for $a_1(2{\cdot}4)$ and $a_1(4)$ is striking in the extreme. Even for $n = 24$ $\sqrt{B_1}\,[a_1(2{\cdot}4)]$ and $B_2\,[a_1(2{\cdot}4)]$ are approaching the values at which a Gram-Charlier approximation to the frequency distribution may be reasonably convergent. Furthermore, the decline in the values of the B's from $n = 24$ to $n = 50$ is marked for $a_1(2{\cdot}4)$, while the decline in the $B[a_1(4)]$ is very slow.

It is accordingly suggested that a table of probability points (perhaps 0·001, 0·01, 0·025, 0·05 and 0·10) of $a(c)$, for c equal to, say, 2·2, be prepared for $n \geqslant 25$ on the assumption that Gram-Charlier applies throughout. For this purpose the values of the mean and variance for n at intervals of, say, 10 should be computed from formulae (4·5) and (4·6); the B_1 and $(B_2 - 3)$ should, however, be computed as for $a_1(c)$. For lower sample sizes it might be well

to use terms to order n^{-2} which would render necessary the use of the fifth and sixth semi-invariants of $a_1(c)$. The formulae given by E. A. Cornish & R. A. Fisher (1937) (assuming Gram-Charlier) could be used to find the probability points. On account of the minuteness of the variance L_2 for c near 2 it will be necessary to work to many places of decimals—at least 10. As stated at the outset, the test of kurtosis $a(2\cdot2)$ will be only slightly less efficient than $a(4)$ and it may be slightly more efficient than $a(1)$, the probability points of which are known approximately for samples of all sizes. In any case the $a(2\cdot2)$ table would be a useful adjunct to that of $a(1)$.

Table 5. *Normal moments, etc., of $a_1(c)$ for $c = 2\cdot4$ and $c = 4$*

	$n=24$		$n=50$	
	$c=2\cdot4$	$c=4$	$c=2\cdot4$	$c=4$
$M_1'=L_1$	1·1662524891	2·769231	1·1721603127	2·884615
$M_2=L_2$	0·001946318	0·559932	0·001058462	0·359550
$M_3=L_3$	0·000069583	0·752488	0·000022251	0·343337
$L_4=M_4-3L_2^2$	0·000004921	1·955999	0·000000919	0·711375
$\sqrt{B_1}=L_3/L_2^{\frac{3}{2}}$	0·8104	1·7960	0·6462	1·5925
$B_2-3=L_4/L_2^2$	1·30	6·24	0·82	5·50

In an earlier paper (1935) the writer suggested that the correlation between b_2 and $a(1)$ for normal samples gave some indication of the relative efficiency of these two tests of normality. In this order of ideas it seems desirable to compute the approximate value of the correlation coefficient between $a(c)$ and $a(c')$, where c and c' are any two positive constants. In the first instance the universe from which the sample of n was drawn was not necessarily normal. Since in the present application we will be concerned only with large samples we assume the universal mean known (and accordingly it may be taken as zero, i.e. $\lambda_1 = 0$), so that, instead of $a(c)$ we use, in reality, $a_1(c)$ given by (3·2). In the remainder of this section we write a for $a_1(c)$ and a' for $a_1(c')$:

$$a = \left(\frac{1}{n}\Sigma\,|x_i|^c\right)\Big/\left(\frac{1}{n}\Sigma x_i^2\right)^{\frac{1}{2}c}, \tag{4·12}$$

$$a' = \left(\frac{1}{n}\Sigma\,|x_i|^{c'}\right)\Big/\left(\frac{1}{n}\Sigma x_i^2\right)^{\frac{1}{2}c'}. \tag{4·13}$$

Set

$$\left.\begin{aligned} y_i &= (|x_i|^c-\mu_{|c|})/\mu_{|c|}, \\ y_i' &= (|x_i|^{c'}-\mu_{|c'|})/\mu_{|c'|}, \\ z_i &= (x_i^2-\mu_2)/\mu_2, \\ \alpha &= \mu_{|c|}/\mu_2^{\frac{1}{2}c}, \quad \alpha' = \mu_{|c'|}/\mu_2^{\frac{1}{2}c'}, \\ C &= \frac{c+c'}{2}, \quad C_k = \frac{C(C+1)(C+2)\dots(C+k-1)}{k!}. \end{aligned}\right\} \tag{4·14}$$

Then

$$\frac{aa'}{\alpha\alpha'} = \left(1+\frac{1}{n}\Sigma y_i\right)\left(1+\frac{1}{n}\Sigma y_i'\right)\left(1+\frac{1}{n}\Sigma z_i\right)^{-\frac{1}{2}(c+c')}. \tag{4·15}$$

The mean value of $aa'/\alpha\alpha'$ was found approximately (i.e. to terms in n^{-3}) by formally expanding the last factor in (4·15), multiplying by the first two factors, and setting down the mean value term by term, so that

$$
\begin{aligned}
M'_{cc'}/\alpha\alpha' = Eaa'/\alpha\alpha' \simeq \Big\{ & 1+\frac{1}{n^2}C_2 nEz^2-\frac{1}{n^3}C_3 nEz^3 \\
& +\frac{1}{n^4}C_4\left(nEz^4+\frac{6n\overline{n-1}}{2}E^2z^2\right)-\frac{1}{n^5}C_5(10n\overline{n-1}\;Ez^3Ez^2) \\
& +\frac{1}{n^6}C_6\,90\,\frac{n\overline{n-1}\,\overline{n-2}}{6}E^3z^2\Big\}+\Big\{-\frac{C_1}{n^2}(nEyz+nEy'z)+\frac{C_2}{n^3}(nEyz^2+nEy'z^2) \\
& -\frac{C'_3}{n^4}[n(Eyz^3+Ey'z^3)+3n\overline{n-1}\;Ez^2(Eyz+Ey'z)] \\
& +\frac{C'_4}{n^5}[4n\overline{n-1}\;Ez^3(Eyz+Ey'z)+6n\overline{n-1}\;Ez^2(Eyz^2+Ey'z^2)] \\
& -\frac{30C'_5}{n^6}\frac{n\overline{n-1}\,\overline{n-2}}{2}E^2z^2(Eyz+Ey'z)\Big\}+\Big\{\frac{1}{n^2}nEyy'-\frac{C'_1}{n^3}nEyy'z \\
& +\frac{C'_2}{n^4}[nEyy'z^2+2n\overline{n-1}\;EyzEy'z+n\overline{n-1}\;Eyy'Ez^2] \\
& -\frac{C_3}{n^5}[n\overline{n-1}\;Eyy'Ez^3+3n\overline{n-1}\;(Eyz^2Ey'z+Ey'z^2Eyz+Eyy'zEz^2)] \\
& +\frac{C'_4}{n^6}\left[\frac{6n\overline{n-1}\,\overline{n-2}}{2}Eyy'E^2z^2+12n\overline{n-1}\,\overline{n-2}\;EyzEy'zEz^2\right]\Big\}.
\end{aligned}
\tag{4·16}
$$

The E's in (4·16) are readily calculable from (4·14), e.g.

$$Eyy' = Ey_iy'_i = E(|x_i|^c-\mu_{|c|})(|x_i|^{c'}-\mu_{|c'|})/\mu_{|c|}\mu_{|c'|} = (\mu_{|c+c'|}/\mu_{|c|}\mu_{|c'|})-1.$$

It has been verified that when c is substituted for c' in (4·13) the formula agrees with that for the second moment of $a_1(c)$ given in § 6.

The coefficient of correlation is, of course,

$$R_{cc'} = M_{cc'}/\surd(M_{cc}M_{c'c'}), \tag{4·17}$$

with

$$M_{cc'} = M'_{cc'}-M'_cM'_{c'}.$$

Formulae for the first and second moments, to the approximation required, for the computation of (4·17) are given in § 6.

As an application, the following are the values of the variances and the covariance for the test of normality $a(1)$ and (b_2), i.e. in which c and c' have respectively the values 1 and 4, and where the universe belongs to the Pearson system with $\lambda_2 = 1$, $\lambda_3 = 0$ and $\lambda_4 = \frac{1}{2}$:

$$
\left.
\begin{aligned}
\frac{M_{cc}}{\mu^2_{|c|}} &\simeq \frac{0·09313705}{n}-\frac{0·262961}{n^2}-\frac{0·196477}{n^3}, \\
\frac{M_{c'c'}}{\mu^2_{|c'|}} &\simeq \frac{4·4286}{n}-\frac{92·25}{n^2}+\frac{831·2}{n^3}, \\
\frac{M_{cc'}}{\mu_{|c|}\mu_{|c'|}} &\simeq -\frac{0·491}{n}+\frac{4·87}{n^2}-\frac{281·5}{n^3}.
\end{aligned}
\right\}
\tag{4·18}
$$

From (4·17) and (4·18), $R_{cc'}(n=100) \doteq -0{\cdot}826$ and $R_{cc'}(n=\infty) = -0{\cdot}764$. It is of great interest to find that, though the universe is markedly non-normal the correlation for indefinitely large samples is practically identical with the normal theory value of $-0{\cdot}767$ (Geary, 1935), another indication, no doubt, that normal theory inferences can usually be applied with confidence when the parent universe is not markedly unsymmetrical.

When samples are indefinitely large we find, from (4·16) and (4·17),

$$R_{cc'} = \frac{4\mu_{|c+c'|} - 2(c\mu_{|c|}\mu_{|c'+2|} + c'\mu_{|c'|}\mu_{|c+2|}) + (cc'\mu_4 - \overline{c-2}\,.\,\overline{c'-2})\,\mu_{|c|}\mu_{|c'|}}{\sqrt{(M_{cc}M_{c'c'})}}, \tag{4·19}$$

where, of course, the values to be taken here for M_{cc} and $M_{c'c'}$ are found by substituting respectively c' for c and c for c' in the numerator. When, in addition, the parent universe is normal, we find

$$R^0_{cc'} = \frac{\left(\frac{c+c'-1}{2}\right)!\sqrt{\pi} - \left(\frac{c-1}{2}\right)!\left(\frac{c'-1}{2}\right)!\left(\frac{cc'+2}{2}\right)}{\sqrt{\left[\left\{\left(\frac{2c-1}{2}\right)!\sqrt{\pi} - \left(\frac{c-1}{2}!\right)^2\left(\frac{c^2+2}{2}\right)\right\}\left\{\left(\frac{2c'-1}{2}\right)!\sqrt{\pi} - \left(\frac{c'-1}{2}!\right)^2\left(\frac{c'^2+2}{2}\right)\right\}\right]}} \tag{4·20}$$

which reduces to $-1/\sqrt{\{12(\pi-3)\}}$ for $c = 1$, $c' = 4$, as it should (Geary, 1935). The following section will accord b_2 (i.e. $a(4)$) a decided primacy amongst tests of normality when the samples are indefinitely large. It may, therefore, be of interest to give the values of the correlation coefficients (for indefinitely large normal samples) between b_2 and $a(c)$ for selected values of c (Table 6). The table suggests, in the high coefficients of correlation, except for c very near 0 or 2, that all the $a(c)$ should be reliable tests of kurtosis, with no great difference between their efficiencies. The efficiency of any two tests would be identical, in the conditions stated, if the coefficient of correlation between them was ± 1 because then, of course, they would be functionally, and not stochastically, related.

Table 6. *Correlation between b_2 and $a(c)$ for indefinitely large normal samples*

Value of c	Value of R^0_{c4}	Value of c	Value of R^0_{c4}
0	0	3	0·980
1	−0·769	4	1
2	0	5	0·983
2·2	0·887	6	0·939
2·5	0·952	∞	0

5. The most efficient tests for indefinitely large samples

In this section we consider the efficiency of tests of kurtosis and asymmetry from the viewpoint of indefinitely large samples.

By definition a test will be regarded as *valid*, in relation to a field of continuous alternative universes including the normal, if its value for infinite samples drawn at random from the normal universe is different from its value for infinite samples from other universes of the field. As the sample number increases the test will become increasingly discriminatory of the normal as distinct from other universes of the field. This increased sensitivity might be given mathematical expression in some such terms as the following: given a probability α (say 0·01), the normal universe W_0 of the field and any other distribution W_1 of the field,

a number n_1 can be found so that for $n \geqslant n_1$ the mean value of the test function for samples of n from W_1 will lie at or beyond the α probability point of the test function for samples of n from W_0: the smaller n_1 the more sensitive the test.

We consider, then, the infinite field of alternative tests of kurtosis represented by (3·1) when c assumes all positive values, and the infinite field of alternative universes represented by the Gram-Charlier frequency

$$\frac{1}{\sqrt{(2\pi)}}\exp\left\{\sum_{i=3}^{\infty}\frac{\lambda_i}{i!}\left(-\frac{d}{dx}\right)^i\right\}e^{-\frac{1}{2}x^2}. \tag{5·1}$$

The universal variance is assumed to be unity, without loss of generality. The normal universe is a member of the field: it is found when all the λ_i ($i > 2$) are zero. We assume that the conditions of § 2 are satisfied so that for indefinitely large samples the frequency distribution of $a(c)$ for all parent universes is normal. Obviously the efficiency of any particular test (i.e. $a(c)$ for a particular value of c) in regard to the normal and a particular non-normal alternative (i.e. a Gram-Charlier frequency with particular values of the λ_i) will be adjudged by considering the ratio of

(i) the difference between the universal mean values of $a(c)$ for the normal and the particular non-normal parent universes; to

(ii) the standard deviation of $a(c)$ for indefinitely large normal samples.

The most efficient test will be $a(c)$ for c a theoretically ascertainable function of the given λ_i which makes the ratio a maximum.

For indefinitely large samples the mean value ϕ of $a(c)$ when the parent universe is given by (5·1) is

$$\phi = \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty}dx\,|x|^c\exp\left\{\sum_i\frac{\lambda_i}{i!}\left(-\frac{d}{dx}\right)^i\right\}e^{-\frac{1}{2}x^2}. \tag{5·2}$$

Obviously
$$\int_{-\infty}^{\infty}dx\,|x|^c\left(-\frac{d}{dx}\right)^{2m+1}e^{-\frac{1}{2}x^2} = 0.$$

Also, when $m \geqslant 1$,

$$\int_{-\infty}^{\infty}dx\,|x|^c\left(\frac{d}{dx}\right)^{2m}e^{-\frac{1}{2}x^2} = \left(\frac{c-1}{2}\right)!\,2^{\frac{1}{2}(c+1)}c(c-2)(c-4)\ldots(c-2m+2), \tag{5·3}$$

a result readily inferable from the obvious fact that the left side vanishes for $c = 0, 2, \ldots, 2m-2$. Accordingly

$$\phi = \left(\frac{c-1}{2}\right)!\frac{2^{\frac{1}{2}(c+1)}}{\sqrt{(2\pi)}}\left\{1+\frac{\lambda_4}{24}c(c-2)+\left(\frac{\lambda_3^2}{72}+\frac{\lambda_6}{720}\right)c(c-2)(c-4)+\ldots\right\}. \tag{5·4}$$

The normal value is given by the first term.

From (4·3), (4·5) and (4·6) it is evident that the value of the standard deviation, for larger normal samples (retaining only $n^{-\frac{1}{2}}$) is

$$\sigma = \frac{2^{\frac{1}{2}c}}{\sqrt{(\pi n)}}\left(\left(\frac{2c-1}{2}\right)!\sqrt{\pi}-\left(\frac{c-1}{2}!\right)^2\frac{c^2+2}{2}\right)^{\frac{1}{2}}. \tag{5·5}$$

The principal term in the deviation $\phi-\phi^0$ (where ϕ^0 is the normal value), from (5·4), is

$$\delta = \frac{\frac{1}{2}(c-1)!\,2^{\frac{1}{2}c}}{\sqrt{\pi}}\cdot\frac{\lambda_4 c(c-2)}{24}. \tag{5·6}$$

To a constant factor, the ratio δ/σ is given by the ***first discriminant***

$$\rho(c) = c(c-2)\left\{\frac{\left(\dfrac{2c-1}{2}\right)!\,\sqrt{\pi}}{\left(\dfrac{c-1}{2}\,!\right)^2} - \frac{c^2+2}{2}\right\}^{-\frac{1}{2}}. \tag{5·7}$$

It will now be shown that $\dfrac{d\rho(c)}{dc} = 0$ for $c = 4$.

The discriminant may be written in the form

$$\rho(c) = c(c-2)\left(\frac{2^c I_{2c}}{I_c} - \frac{c^2+2}{2}\right)^{-\frac{1}{2}}, \tag{5·8}$$

where

$$I_c = \int_0^{\frac{1}{2}\pi} \cos^c\theta\, d\theta, \tag{5·9}$$

and

$$\frac{\rho'(c)}{\rho(c)} = \frac{1}{c} + \frac{1}{c-2} - \frac{1}{2}\left\{2^c\left(\frac{I_{2c}\log 2}{I_c} + \frac{I'_{2c}}{I_c} - \frac{I_{2c}I'_c}{I_c^2}\right) - c\right\}\Big/\left\{\frac{2^c I_{2c}}{I_c} - \frac{c^2+2}{2}\right\}. \tag{5·10}$$

From (5·9)

$$J_c = I'_c = \int_0^{\frac{1}{2}\pi} d\theta \log^c\theta \log\cos\theta. \tag{5·11}$$

From a fairly well-known property

$$J_0 = \int_0^{\frac{1}{2}\pi} d\theta \log\cos\theta = -\tfrac{1}{2}\pi\log 2. \tag{5·12}$$

In (5·10) we shall be concerned only with even positive integer values of c. We have at once

$$I_0 = \frac{\pi}{2},\quad I_2 = \frac{\pi}{4},\quad I_4 = \frac{3\pi}{16},\quad I_6 = \frac{5\pi}{32},\quad I_8 = \frac{35\pi}{256}. \tag{5·13}$$

From (5·11) $\quad J_{2c} = \displaystyle\int_0^{\frac{1}{2}\pi} d\theta \cos^{2c}\theta \log\cos\theta = \int_0^{\frac{1}{2}\pi} d(\sin\theta)\cos^{2c-1}\theta\log\cos\theta,$

which, by partial integration,

$$= \int_0^{\frac{1}{2}\pi} d\theta \sin\theta\left(\overline{2c-1}\sin\theta\cos^{2c-2}\theta\log\cos\theta + \frac{\cos^{2c-1}\theta\sin\theta}{\cos\theta}\right)$$

$$= (2c-1)(J_{2c-2} - J_{2c}) + I_{2c-2} - I_{2c}.$$

Hence

$$2cJ_{2c} = (2c-1)J_{2c-2} - I_{2c} + I_{2c-2}. \tag{5·14}$$

From (5·12), (5·13) and (5·14),

$$\left.\begin{aligned} J_0 &= -\tfrac{1}{2}\pi\log 2, & J_6 &= (-60\pi\log 2 + 37\pi)/384,\\ J_2 &= (-2\pi\log 2 + \pi)/8, & J_8 &= (-840\pi\log 2 + 533\pi)/6144.\\ J_4 &= (-12\pi\log 2 + 7\pi)/64, & & \end{aligned}\right\} \tag{5·15}$$

Noting that $I'_{2c} = 2J_{2c}$ and substituting in the right side of (5·10) the values of I and J given by (5·13) and (5·15), we find $\rho'(4) = 0$. Table 7 gives the values of the discriminant for certain values of c.

The discriminant accordingly assumes a maximum value for $c = 4$, a result so remarkable that one might be inclined to suspect that it is a consequence of the form which was assumed for the alternative to the normal curve, a form which, in placing such emphasis on λ_4

high-lights, so to speak, b_2 ($= \lambda_4 + 3$ when $\lambda_2 = 1$ for indefinitely large samples) as a test of normality. From the algebraic point of view this is anything but obvious: the property emerges from quite a complicated piece of algebra. It may also be emphasized that the field of alternatives (5·1) is not arbitrary; it is a general form of frequency distribution when all the λ_i are finite. Admittedly the discriminant takes account only of the term in λ_4 in the expansion; but this is certainly the most significant term for a wide class of frequency distributions, namely, those of homogeneous symmetrical functions of samples of n as n tends towards infinity under very general conditions for the parent universe, provided that the resulting frequency distribution can be assumed to have its third moment zero; for then the only term in n^{-1} in the frequency distribution of the function will be the term in λ_4. The significance of the property demonstrated must not be overstressed since it is subject to many qualifications, but it gives strong grounds for holding that, for very large samples, b_2 is the most efficient test of normality of tests of type $a(c)$ in relation to a very extended class of alternative universes. At the same time Table 7 shows that there can be little difference in efficiency in the field $a(c)$ for c ranging from close to 2 to about 5. There is but little doubt, on this showing, that b_2 is more sensitive than $a(1)$, a conclusion suggested on the basis of certain experimental results by E. S. Pearson (1935) and examined from the viewpoint of power function theory in § 6.

Table 7

$0<c<2$	Discriminant $\rho(c)$	$2<c<\infty$	Discriminant $\rho(c)$
+0	−2·334	2+0	4·460
0·1	−2·541	2·1	4·508
0·2	−2·725	2·5	4·666
0·5	−3·188	3·0	4·801
0·7	−3·441	3·9	4·898
1·0	−3·758	**4·0**	**4·900**
1·1	−3·851	4·1	4·898
1·5	−4·166	5·0	4·818
1·9	−4·405	6·0	4·602
2−0	−4·460	7·0	4·288
		8·0	3·906

Adverting to (5·4) in conjunction with (5·5), it might be asked if, on the analogy of the maximal property just demonstrated for the first discriminant, the function

$$\rho_2(c) = c(c-2)(c-4)\left\{\frac{\left(\frac{2c-1}{2}\right)!\sqrt{\pi}}{\left(\frac{c-1}{2}!\right)^2} - \frac{c^2+2}{2}\right\}^{-\frac{1}{2}}$$

has a turning point at $c = 6$. The answer is in the negative. The value of $\rho_2'(6)/\rho_2(6)$ is, in fact, 15/34. At the same time there must be a zero of $\rho_2'(c)$ very near $c = 6$ since

$$\rho_2(5\cdot 9) = 8\cdot 79, \quad \rho_2(6) = 9\cdot 20, \quad \rho_2(6\cdot 1) = 8\cdot 56.$$

Analogous to the field on tests of kurtosis represented by (3·1) we may consider as a field of tests of asymmetry:

$$g(c) = \frac{1}{n}\{-\Sigma' \mid x_i - \bar{x} \mid^c + \Sigma''(x_i - \bar{x})^c\} \Big/ \left\{\frac{1}{n}\Sigma(x_i-\bar{x})^2\right\}^{\frac{1}{2}c}, \qquad (5\cdot 16)$$

where Σ' extends to the observations x_i less than the mean $\bar{x}$ and Σ'' to the rest of the sample. For $c = 3$ the test is, of course, $\sqrt{b_1}$. For normal samples

$$E\{g(c)\}^k = E\left\{-\frac{1}{n}\Sigma' \mid x_i - \bar{x} \mid^c + \frac{1}{n}\Sigma''(x_i - \bar{x})^c\right\}^k \Big/ E\left\{\frac{1}{n}\Sigma(x_i - \bar{x})^2\right\}^{\frac{1}{2}kc}, \tag{5·17}$$

the denominator of which is identical with the denominator of (4·1). Knowing the joint distribution (for normal samples) of $(x_1 - \bar{x})$, $(x_2 - \bar{x})$, ... (Geary, 1936), there is no theoretical difficulty in finding the mean values of the terms of the numerator for positive integer values of k. Here we shall be concerned only with the first and second moments, i.e. those for (5·17) for $k = 1$ and $k = 2$. We require the normal distribution of $z_1 = x_1 - \bar{x}$ and the joint distribution of z_1 and $z_2 = x_2 - \bar{x}$. These are

$$(z_1):\quad \left(\frac{n}{2\pi\, n-1}\right)^{\frac{1}{2}} \exp\left\{-\frac{nz_1^2}{2(n-1)}\right\} dz_1,$$

$$(z_1, z_2):\quad \frac{1}{2\pi}\left(\frac{n}{n-2}\right)^{\frac{1}{2}} \exp\left\{-\frac{(n-1)(z_1^2+z_2^2)}{2(n-2)} - \frac{z_1 z_2}{(n-2)}\right\} dz_1 dz_2 = f(z_1, z_2)\, dz_1 dz_2. \tag{5·18}$$

Clearly the odd normal moments of $g(c)$ are zero. Then

$$E\left\{-\frac{1}{n}\Sigma' \mid x_i - \bar{x} \mid^c + \frac{1}{n}\Sigma''(x_i - \bar{x})^c\right\}^2 = \frac{n}{n^2} E \mid z_1 \mid^{2c} + \frac{n(n-1)}{n^2} E_1(z_1, z_2), \tag{5·19}$$

where $E_1(z_1, z_2)$ is the mean value of the two-dimensional terms. We then have

$$\begin{aligned} E_1(z_1, z_2) &= \int_{-\infty}^{0} (-z_1)^c dz_1 \int_{-\infty}^{0} (-z_2)^c dz_2 f(z_1, z_2) - \int_{-\infty}^{0} dz_1 (-z_1)^c \int_0^\infty dz_2 z_2^c f(z_1, z_2) \\ &\quad - \int_0^\infty dz_1 z_1^c \int_{-\infty}^{0} dz_2 (-z_2)^c f(z_1, z_2) + \int_0^\infty dz_1 z_1^c \int_0^\infty dz_2 z_2^c f(z_1, z_2) \\ &= \int_0^\infty \int_0^\infty z_1^c z_2^c dz_1 dz_2 \{f(-z_1, -z_2) - f(-z_1, z_2) - f(z_1, -z_2) + f(z_1, z_2)\} \\ &= -\frac{2^{c+2}}{2\pi}\left(\frac{n}{n-2}\right)^{\frac{1}{2}} \frac{(n-2)^{c+1}}{(n-1)^{c+2}} \left(\frac{c}{2}!\right)^2 \left\{1 + \frac{(c+2)^2}{3!(n-1)^2} + \frac{(c+2)^2 (c+4)^2}{5!(n-1)^4} + \ldots\right\}, \end{aligned} \tag{5·20}$$

$$Ez_1^{2c} = \frac{2c-1}{2}!\left(\frac{2n-1}{n}\right)^c \frac{1}{\sqrt{\pi}}. \tag{5·21}$$

Also

$$E\left\{\frac{1}{n}\Sigma(x_i - \bar{x})^2\right\}^c = \left(\frac{2}{n}\right)^c \left(\frac{n+2c-3}{2}\right)! \Big/ \left(\frac{n-3}{2}\right)!. \tag{5·22}$$

We now have all the expressions required for the variance of normal $g(c)$. We require, for what follows, only the term in n^{-1} which is

$$\sigma^2 = \frac{1}{n}\left\{\left(\frac{2c-1}{2}\right)! \frac{2^c}{\sqrt{\pi}} - \left(\frac{c}{2}!\right)^2 \frac{2^{c+1}}{\pi}\right\}. \tag{5·23}$$

Consider now a field of alternative universes represented by

$$\frac{1}{\sqrt{(2\pi)}}\left\{1 + \frac{\lambda_3}{6}(x^3 - 3x)\right\} e^{-\frac{1}{2}x^2}, \tag{5·24}$$

the 'first approximation to the law of error' (for universal variance unity), obviously the most appropriate asymmetrical field, for different values of the parameter λ_3, and con-

taining as a member of the field the normal distribution found for $\lambda_3 = 0$. For indefinitely large samples from (5·24) the mean value of $g(c)$ is

$$\delta = \frac{2\lambda_3}{6\sqrt{(2\pi)}}\int_0^\infty dx\, x^c(x^3-3x)\,e^{-\frac{1}{2}x^2} = \frac{c}{2}!\,(c-1)\,2^{\frac{1}{2}c}\frac{\lambda_3}{3}\sqrt{\frac{1}{2\pi}}. \tag{5·25}$$

From (5·23) and (5·25)

$$\frac{\delta}{\sigma} = \frac{\lambda_3 n^{\frac{1}{2}}}{6}\tau(c), \tag{5·26}$$

the *skew discriminant* $\tau(c)$ being given by

$$\tau(c) = (c-1)\left\{\left(\frac{2c-1}{2}\right)!\left(\frac{c}{2}!\right)^{-2}\frac{\sqrt{\pi}}{2}-1\right\}^{-\frac{1}{2}} = (c-1)\left(\frac{2^{c+1}}{2c+1}\frac{I_{2c+2}}{I_{c+1}}-1\right)^{-\frac{1}{2}}. \tag{5·27}$$

Log-differentiating,

$$\frac{\tau'(c)}{\tau(c)} = \frac{1}{c-1} - \frac{2^{c+1}}{2}\left\{\frac{1}{2c+1}\left(\frac{2J_{2c+2}}{I_{c+1}} - \frac{I_{2c+2}J_{c+1}}{I_{c+1}^2}\right)\right.$$
$$\left. - \frac{I_{2c+2}}{I_{c+1}}\frac{2}{(2c+1)^2} + \frac{I_{2c+2}}{I_{c+1}}\frac{\log 2}{2c+1}\right\}\left\{\frac{2^{c+1}}{2c+1}\frac{I_{2c+2}}{I_{c+1}}-1\right\}^{-1}. \tag{5·28}$$

Setting $c = 3$ and using (5·13) and (5·15), we find that $\tau'(3) = 0$. Values of $\tau(c)$ for four values of c are as follows:

c	$\tau(c)$	c	$\tau(c)$
2	2·370	4	2·389
3	2·450	5	2·236

Accordingly, for indefinitely large samples the test of asymmetry $g(c)$ is most efficient for $c = 3$, when the test becomes the familiar $\sqrt{b_1}$. The margin in favour of this value of c, as compared with others in the range $2 \leqslant c \leqslant 5$, is, however, quite small.

6. TESTS OF KURTOSIS FROM THE POWER FUNCTION VIEWPOINT

It may be useful to open this section with an interpretation of the results of the previous section from the point of view of the type of error theory of J. Neyman & E. S. Pearson (1933, 1936). For this we consider two universes of the field, the normal W_0 and any non-normal universe W_1, and two tests of kurtosis $a(4) = b_2$ and $a(c_1)$ for a particular value c_1 of c. Suppose that samples are sufficiently large that $a(c)$, for samples from all universes of the field, may be regarded as normally distributed.

Given a probability α, a sample number n can be found so that the mean value of $a(c_1)$ from W_1 lies exactly at, say, the upper α probability point of the distribution of $a(c_1)$ from W_0. Then from the results established in the preceding section the value of $a(4)$ for the same sample of n from W_1 could lie beyond the α probability point of $a(4)$ for normal samples of n. Suppose that the rule adopted was to regard as non-normal all samples for which $a(c)$ lies beyond the normal α probability point, and suppose that a very large number N of samples were drawn, N_0 from universes not significantly different from normal (defining 'insignificance' in some manner) and N_1 from non-normal universes, so that $N = N_0 + N_1$, where N_0 and N_1 are not necessarily known in advance. Then using $a(c_1)$ the number of erroneous allocations will be approximately $\alpha N_0 + \frac{1}{2}N_1$, whereas using $a(4)$ the number will be $\alpha N_0 + (\frac{1}{2}-p)\,N_1$ $(\frac{1}{2} > p > 0)$, showing a definite advantage in favour of $a(4)$. The same conclusion emerges whatever value of $c \neq 4$ or whatever non-normal universe be taken for comparison.

The type of error approach reveals the theoretical weakness of using the method of § 5 for the assessment of relative efficiency of tests of normality; namely that the proportion of

errors of judgment, even using $a(4)$, remains large, due fundamentally to concentrating on a single value (the mean) as typical or representative of samples from the non-normal universe; it is also a disadvantage that the sample number n_1 is necessarily a function of the particular value c_1 of c. The method has further disadvantages of which the principal are perhaps (i) a somewhat restricted field of alternative universes; (ii) the assumption that the samples were indefinitely large, essential to justify the normality of $a(c)$ for samples from any member of the universe field.

The Neyman-Pearson power function approach which will now be considered cannot be regarded as entirely free from these objections in its application to the material so far available from this research. It enables us, at any rate, to contemplate samples which, if not small, are within the range of experimental practicability.

The problem of the relative efficiency of the different members of a field of tests of kurtosis $a(c)$ will now be considered in its power function aspects. For the present purpose the *power* may be defined as follows:

Given a probability α (say 0·01), a sample number n, a particular value c_1 of c and a non-normal parent universe W_1, the power, in relation to these data, represents the frequency of $a(c_1)$ for samples drawn at random from W_1 lying beyond the α probability point for $a(c_1)$ computed from samples drawn from a normal universe. The greater the power the more discriminatory the test. Accordingly, it is in theory necessary to know the frequency distribution of $a(c)$ for all sample sizes, for all values of c and for all universes. Considering that the only frequency distribution of the field contemplated which can be regarded as determined for all sample sizes is $a(1)$ for normal samples (Geary, 1935, 1936), many compromises are necessary to give any kind of practical effect to the power concept. The compromises proposed are as follows:

(1) The form $a_1(c)$, given by (3·2), is used instead of the form $a(c)$ given by (3·1).

(2) Only large samples are dealt with.

(3) The field of alternative universes is restricted.

Using $a_1(c)$, the first four moments (from the origin) of $a_1(c)$ for samples from any universe can be expanded without real difficulty, and so approximate frequency distributions (using the Karl Pearson or Gram-Charlier systems) can be obtained. As to (1), from experiments in $a(1)$ and $a(4)$ the writer has verified that, for medium-sized normal samples, there is little difference between the probability points (e.g. 0·01, 0·05) of $a_1(c)$ and $a(c)$, though the higher semi-invariants (given n) are larger for the latter. In regard to (2) and (3) little confidence could be reposed in the values of the moments computed from expansions even to n^{-3} unless the sample number was at least of the order of 100 when c is greater than, say, 3; and, even if the moments were known exactly, the empirical frequencies would be more than doubtful for small samples. The approach finds its main justification in the consideration that any errors due to these necessary compromises may be presumed to apply more or less equally and in the same direction to the tests of kurtosis compared; generous, perhaps too generous, advantage is taken of this justification in the concluding part of this section.

Set, then,
$$a_1(c) = \left\{\frac{1}{n}\Sigma\,|x_i|^c\right\}\Big/\left\{\frac{1}{n}\Sigma x_i^2\right\}^{\frac{1}{2}c}, \tag{6·1}$$

so that
$$\frac{a_1(c)}{\alpha} = \left(1+\frac{1}{n}\Sigma y_i\right)\left(1+\frac{1}{n}\Sigma z_i\right)^{-\frac{1}{2}c}, \tag{6·2}$$

where
$$\alpha = \mu_{|c|}/\mu_2^{\frac{1}{2}c}, \quad y_i = (|x_i|^c-\mu_{|c|})/\mu_{|c|}, \quad z_i = (x_i^2-\mu_2)/\mu_2, \tag{6·3}$$

the universal mean being taken as zero, without loss of generality. Raising (6·2) to powers 1, 2, 3, 4, expanding to the required degree the final factor, multiplying by the first factor on the right, and setting down the mean value of each term we find, to n^{-3},

$$
\begin{aligned}
M'_1/\alpha = {} & 1-\frac{1}{n}\{k_1^{(1)}(11)-k_2^{(1)}(02)\}+\frac{1}{n^2}\{k_2^{(1)}(12)-k_3^{(1)}[(03)+3(11)(02)]+3k_4^{(1)}(02)^2\} \\
& +\frac{1}{n^3}\{k_3^{(1)}[3(11)(02)-(13)]+k_4^{(1)}[(04)-3(02)^2+4(11)(03)+6(12)(02)] \\
& -k_5^{(1)}[10(03)(02)+15(11)(02)^2]+15k_6^{(1)}(02)\},
\end{aligned}
\tag{6·4}
$$

$$
\begin{aligned}
M'_2/\alpha^2 = {} & 1+\frac{1}{n}\{k_2^{(2)}(02)-2k_1^{(2)}(11)+(20)\}+\frac{1}{n^2}\{-k_3^{(2)}(03)+3k_4^{(2)}(02)^2 \\
& +2k_2^{(2)}(12)-6k_3^{(2)}(11)(02)-k_1^{(2)}(21)+k_2^{(2)}(20)(02)+2k_2^{(2)}(11)^2\} \\
& +\frac{1}{n^3}\{k_4^{(2)}[(04)-3(02)^2]-10k_5^{(2)}(03)(02)+15k_6^{(2)}(02)^3-2k_3^{(2)}[(13)-3(11)(02)] \\
& +4k_4^{(2)}[2(11)(03)+3(12)(02)]-30k_5^{(2)}(11)(02)^2 \\
& +k_2^{(2)}[(22)-(20)(02)]-k_3^{(2)}[(20)(03)+3(21)(02)]-2k_2^{(2)}(11)^2 \\
& -6k_3^{(2)}(12)(11)+12k_4^{(2)}(11)^2(02)+3k_4^{(2)}(20)(02)^2\},
\end{aligned}
\tag{6·5}
$$

$$
\begin{aligned}
M'_3/\alpha^3 = {} & 1+\frac{1}{n}\{k_2^{(3)}(02)-3k_1^{(3)}(11)+3(20)\}+\frac{1}{n^2}\{-k_3^{(3)}(03)+3k_4^{(3)}(02)^2 \\
& +3k_2^{(3)}(12)-9k_3^{(3)}(11)(02)-3k_1^{(3)}(21)+3k_2^{(3)}(20)(02)+6k_2^{(3)}(11)^2 \\
& +(30)-3k_1^{(3)}(20)(11)\}+\frac{1}{n^3}\{k_4^{(3)}[(04)-3(02)^2]-10k_5^{(3)}(03)(02)+15k_6^{(3)}(02)^3 \\
& -3k_3^{(3)}[(13)-3(11)(02)]+6k_4^{(3)}[2(11)(03)+3(12)(02)]-45k_5^{(3)}(11)(02)^2 \\
& +3k_2^{(3)}[(22)-(20)(02)]-3k_3^{(3)}[(20)(03)+3(21)(02)]+9k_4^{(3)}(20)(02)^2 \\
& -6k_2^{(3)}(11)^2-18k_3^{(3)}(12)(11)+36k_4^{(3)}(11)^2(02)-k_1^{(3)}(31) \\
& +k_2^{(3)}(30)(02)+3k_1^{(3)}(20)(11) \\
& +3k_2^{(3)}[(20)(12)+2(21)(11)]-9k_3^{(3)}(20)(11)(02)-6k_3^{(3)}(11)^3\},
\end{aligned}
\tag{6·6}
$$

$$
\begin{aligned}
M'_4/\alpha^4 = {} & 1+\frac{1}{n}\{k_2^{(4)}(02)-4k_1^{(4)}(11)+6(20)\}+\frac{1}{n^2}\{-k_3^{(4)}(03)+3k_4^{(4)}(02)^2 \\
& +4k_2^{(4)}(12)-12k_3^{(4)}(11)(02)-6k_1^{(4)}(21)+6k_2^{(4)}(20)(02)+12k_2^{(4)}(11)^2 \\
& +4(30)-12k_1^{(4)}(20)(11)+3(20)^2\}+\frac{1}{n^3}\{k_4^{(4)}[(04)-3(02)^2] \\
& -10k_5^{(4)}(03)(02)-15k_6^{(4)}(02)^3-4k_3^{(4)}[(13)-3(11)(02)] \\
& +8k_4^{(4)}[2(11)(03)+3(12)(02)]-60k_5^{(4)}(11)(02)^2+6k_2^{(4)}[(22)-(20)(02)] \\
& -6k_3^{(4)}[(20)(03)+3(21)(02)]+18k_4^{(4)}(20)(02)^2-12k_2^{(4)}(11)^2 \\
& -36k_3^{(4)}(12)(11)+72k_4^{(4)}(11)^2(02)-4k_1^{(4)}(31)+4k_2^{(4)}(30)(02) \\
& +12k_1^{(4)}(20)(11)+12k_2^{(4)}[(12)(20)+2(21)(11)] \\
& -36k_3^{(4)}(20)(11)(02)-24k_3^{(4)}(11)^3+(40)-4k_1^{(4)}(30)(11)-3(20)^2 \\
& -6k_1^{(4)}(20)(21)+3k_2^{(4)}(20)^2(02)+12k_2^{(4)}(20)(11)^2\},
\end{aligned}
\tag{6·7}
$$

where $$k_r^{(p)}=\frac{\tfrac{1}{2}pc(\tfrac{1}{2}pc+1)(\tfrac{1}{2}pc+2)\dots(\tfrac{1}{2}pc+r-1)}{r!},\quad (fg)=Ey_i^f z_i^g,$$

the latter, of course, the same for all i. The (fg) required for the computation of (6·4)–(6·7) are

$$\left.\begin{aligned}
(11) &= (\mu_{|2+c|} - \mu_2\mu_{|c|})/\mu_2\mu_{|c|},\\
(02) &= (\mu_4 - \mu_2^2)/\mu_2^2,\\
(12) &= (\mu_{|4+c|} - 2\mu_{|2+c|}\mu_2 - \mu_{|c|}\mu_4 + 2\mu_{|c|}\mu_2^2)/\mu_{|c|}\mu_2^2,\\
(03) &= (\mu_6 - 3\mu_4\mu_2 + 2\mu_2^3)/\mu_2^3,\\
(04) &= (\mu_8 - 4\mu_6\mu_2 + 6\mu_4\mu_2^2 - 3\mu_2^4)/\mu_2^4,\\
(13) &= [\mu_{|6+c|} - 3\mu_{|4+c|}\mu_2 + 3\mu_{|2+c|}\mu_2^2 - \mu_{|c|}(\mu_6 - 3\mu_4\mu_2 + 3\mu_2^3)]/\mu_{|c|}\mu_2^3,\\
(21) &= [\mu_{|2c+2|} - 2\mu_{|c+2|}\mu_{|c|} - \mu_2(\mu_{|2c|} - 2\mu_{|c|}^2)]/\mu_{|c|}^2\mu_2,\\
(22) &= (\mu_{|2c+4|} - 2\mu_{|2c+2|}\mu_2 + \mu_{|2c|}\mu_2^2 - 2\mu_{|c+4|}\mu_{|c|} + 4\mu_{|c+2|}\mu_{|c|}\mu_2 - 3\mu_{|c|}^2\mu_2^2 + \mu_{|c|}^2\mu_4)/\mu_{|c|}^2\mu_2^2,\\
(20) &= (\mu_{|2c|} - \mu_{|c|}^2)/\mu_{|c|}^2,\\
(30) &= (\mu_{|3c|} - 3\mu_{|2c|}\mu_{|c|} + 2\mu_{|c|}^3)/\mu_{|c|}^3,\\
(31) &= (\mu_{|3c+2|} - 3\mu_{|2c+2|}\mu_{|c|} + 3\mu_{|c+2|}\mu_{|c|}^2 - \mu_{|3c|}\mu_2 + 3\mu_{|2c|}\mu_{|c|}\mu_2 - 3\mu_{|c|}^3\mu_2)/\mu_{|c|}^3\mu_2,\\
(40) &= (\mu_{|4c|} - 4\mu_{|3c|}\mu_{|c|} + 6\mu_{|2c|}\mu_{|c|}^2 - 3\mu_{|c|}^4)/\mu_{|c|}^4.
\end{aligned}\right\} \tag{6·8}$$

(6·8) is, of course, an immediate consequence of (6·3). The writer has checked the accuracy of formulae (6·4)–(6·7) by reference to the normal universe for $c = 1$.

The reader will have no illusions as to the magnitude of the task of applying the foregoing theory to particular cases. The formulae are set down, however, in the hope that other researchers will be sufficiently sensible of the importance of the theory to assist in building up a fairly extensive set of results. The writer has to be content, in the meantime, to consider the case of the symmetrical universe field given by

$$\frac{1}{\sqrt{(2\pi)}}\left\{1 + \frac{\lambda_4}{24}\left(\frac{d}{dx}\right)^4\right\} e^{-\frac{1}{2}x^2}. \tag{6·9}$$

when $\lambda_4 = \frac{1}{2}$, the normal being given, of course, for $\lambda_4 = 0$, and for $c = 4$ and $c = 1$. These values of c are selected because the theory in § 5 has suggested that $a(4)$ is probably the most efficient of the test-field $a(c)$, while $a(1)$ is the only member of the field for which the normal

Table 8. *Moments from formulae* (6·8)

(fg)	$c=4$		$c=1$	
	Normal	$\lambda_4=\frac{1}{2}$	Normal	$\lambda_4=\frac{1}{2}$
(11)	4	5·428571	1	1·17021276
(02)	2	2·5	2	2·5
(12)	24	45·64286	3	4·88297871
(03)	8	14	8	14
(04)	60	138	60	138
(13)	216	544·2857	21	44·106383
(21)	256/3	177·71428	1·141593	1·75544898
(22)	2,720/3	2,481·92857	7·707963	14·766814
(20)	32/3	16·142857	0·570796	0·63834981
(30)	352	799·142857	0·429204	0·6405182
(31)	4,352	12,785·2653	3	5·236134
(40)	23,552	73,250·178	—	2·002492

distribution is known for samples of all sizes. The necessary moments (fg) given by (6·8) are shown in Table 8. Based on the values in this table, moments (M') given by (6·4)–(6·7) of $a_1(c)$ and semi-invariants (L) derived therefrom are as follows. The normal values are, of course, known exactly but were computed for the purpose of checking the formulae:

$c = 4$; *normal universe*

$$\frac{L_1}{3} = \frac{M_1'}{3} \simeq 1 - \frac{2}{n} + \frac{4}{n^2} - \frac{8}{n^3},$$

$$\frac{M_2'}{9} \simeq 1 - \frac{4}{3n} - \frac{28}{n^2} + \frac{1040}{3n^3}, \quad \frac{L_2}{9} \simeq \frac{8}{3n} - \frac{40}{n^2} + \frac{1136}{3n^3},$$

$$\frac{M_3'}{27} \simeq 1 + \frac{2}{n} - \frac{48}{n^2} - \frac{1040}{n^3}, \quad \frac{L_3}{27} \simeq \frac{64}{n^2} - \frac{2368}{n^3},$$

$$\frac{M_4'}{81} \simeq 1 + \frac{8}{n} + \frac{40}{3n^2} - \frac{3520}{n^3}, \quad \frac{L_4}{81} \simeq \frac{3840}{n^3}.$$

$c = 4$; *universal* $\lambda_4 = \frac{1}{2}$

$$\frac{L_1}{3\cdot5} = \frac{M_1'}{3\cdot5} \simeq 1 - \frac{3\cdot357}{n} + \frac{11\cdot822}{n^2} + \frac{12\cdot1}{n^3},$$

$$\frac{M_2'}{(3\cdot5)^2} \simeq 1 - \frac{2\cdot286}{n} - \frac{57\cdot34}{n^2} + \frac{776\cdot03}{n^3}, \quad \frac{L_2}{(3\cdot5)^2} \simeq \frac{4\cdot4286}{n} - \frac{92\cdot25}{n^2} + \frac{831\cdot2}{n^3},$$

$$\frac{M_3'}{(3\cdot5)^3} \simeq 1 + \frac{3\cdot215}{n} - \frac{107\cdot47}{n^2} - \frac{2853\cdot89}{n^3}, \quad \frac{L_3}{(3\cdot5)^3} \simeq \frac{144\cdot61}{n^2} - \frac{6193\cdot95}{n^3},$$

$$\frac{M_4'}{(3\cdot5)^4} \simeq 1 + \frac{13\cdot143}{n} + \frac{20\cdot49}{n^2} - \frac{9529}{n^3}, \quad \frac{L_4}{(3\cdot5)^4} \simeq \frac{10{,}587}{n^3}.$$

$c = 1$; *normal universe*

$$L_1 = M_1' \simeq 0\cdot7978845608 + \frac{0\cdot19947114}{n} + \frac{0\cdot02493389}{n^2} - \frac{0\cdot03116737}{n^3},$$

$$L_2 \simeq \frac{0\cdot04507034}{n} - \frac{0\cdot07957747}{n^2} + \frac{0\cdot03978874}{n^3},$$

$$L_3 \simeq -\frac{0\cdot01685645}{n^2} + \frac{0\cdot07613597}{n^3}.$$

$c = 1$; *universal* $\lambda_4 = \frac{1}{2}$

$$\frac{L_1}{\mu_{|1|}} = \frac{M_1'}{\mu_{|1|}} \simeq 1 + \frac{0\cdot35239362}{n} - \frac{0\cdot159616}{n^2} - \frac{0\cdot745838}{n^3},$$

$$\frac{M_2'}{\mu_{|1|}^2} \simeq 1 + \frac{0\cdot79792429}{n} - \frac{0\cdot458012}{n^2} - \frac{1\cdot800648}{n^3}, \quad \frac{L_2}{\mu_{|1|}^2} \simeq \frac{0\cdot09313705}{n} - \frac{0\cdot262961}{n^2} - \frac{0\cdot196477}{n^3},$$

$$\frac{M_3'}{\mu_{|1|}^3} \simeq 1 + \frac{1\cdot336592}{n} - \frac{0\cdot850081}{n^2} - \frac{3\cdot239101}{n^3}, \quad \frac{L_3}{\mu_{|1|}^3} \simeq \frac{0\cdot053356}{n^2} + \frac{0\cdot204164}{n^3},$$

$$\mu_{|1|} = 0\cdot78126197.$$

Two sample sizes were considered: $n = 100$ and $n = 500$. For $n = 100$ and $c = 4$, the

following are the Pearson Type IV frequencies of $a_1(4)$ when the parent universes are normal and have $\lambda_4 = \beta_2 - 3 = \frac{1}{2}$ respectively:

$$\text{Normal: } \lambda_4 = 0. \quad \left.\begin{array}{l} \kappa \cos^{11\cdot 3350}\theta\, e^{13\cdot 015438\theta}\, dx, \\ \tan\theta = (x - 1\cdot 873387)/0\cdot 765849, \\ \log_{10}\kappa = \bar{3}\cdot 2644596. \end{array}\right\} \tag{6·10}$$

$$\lambda_4 = \tfrac{1}{2}: \quad \left.\begin{array}{l} \kappa \cos^{6\cdot 0096}\theta\, e^{2\cdot 31289\theta}\, dx, \\ \tan\theta = (x - 2\cdot 8522)/0\cdot 9062, \\ \log_{10}\kappa = \bar{1}\cdot 7499974. \end{array}\right\} \tag{6·11}$$

The normal probability points shown in column (2) of Table 10 were derived from the foregoing normal frequency (6·10); the points in column (3) were derived from a Gram-Charlier formula (Geary, 1935). The 0·01 and 0·05 points given in column (2) are practically identical with those given by E. S. Pearson (1929) for $a(4)$, namely, 4·39 and 3·77. The powers given in column (4) are the aggregate frequencies lying beyond the values of the variate shown in column (2) on the assumption that the actual frequency was (6·11). The corresponding figures for $c = 1$ given in column (5) were based on a Gram-Charlier formula.

Table 9. *Power of $a_1(c)$ for $c = 4$ and $c = 1$ of discriminating* (6·9) *for $\lambda_4 = \frac{1}{2}$ from the normal ($\lambda_4 = 0$) at four normal theory probability levels. Samples of* 100

Normal theory probability	Normal theory probability points		Power for frequency (6·9) with $\lambda_4 = \frac{1}{2}$	
	$c = 4$ (upper)	$c = 1$ (lower)	$c = 4$	$c = 1$
(1)	(2)	(3)	(4)	(5)
0·01	4·3836	0·7482	0·0648	0·0695
0·05	3·7744	0·7642	0·1995	0·1979
0·10	3·5195	0·7725	0·3163	0·3037
0·20	3·3110	0·7824	0·4525	0·4597

Before discussing the comparative powers in Table 9 it will be convenient to give a table, 11, on the same lines but for $n = 500$. On account of the larger sample size it has been necessary to change the reference-probabilities given in column (1). For the construction of this table Gram-Charlier formulae were used throughout—the probability points being determined from the E. A. Cornish & R. A. Fisher (1937) formulae—after verifying that for two of the probability levels, 0·01 and 0·05, the probability points for $c = 4$ (column (2) above) did not differ appreciably from those given by E. S. Pearson, namely, 3·60 and 3·37 (for $a(4)$), based on a Type IV curve.

The analysis in § 5 has enabled us to come fairly firmly to the conclusion that for indefinitely large samples $a(4)$ was to be preferred to $a(1)$ as a test of normality. We see from Tables 9 and 10 that this is subject to an important qualification. Table 9 shows that the discriminating power is definitely greater for samples of 500 for $a(4)$ than for $a(1)$, but the superiority is less emphatic than might have been anticipated from § 5. For medium-sized samples (Table 9) $a(4)$ exhibits no superiority. Of course, these conclusions are very tentative, as being based upon a single alternative and on particular sample sizes. The writer had proposed, in addition, to examine the universes (i) $\lambda_3 = 0$, $\lambda_4 = 1$ and (ii) $\lambda_3^2 = \lambda_4 = \frac{1}{2}$ as alternatives to the normal but time did not permit; he ventures to repeat the hope that other students will take the matter up.

Table 10. *Power of $a_1(c)$ for $c = 4$ and $c = 1$ of discriminating (6·9) for $\lambda_4 = \frac{1}{2}$ from the normal ($\lambda_4 = 0$) at four probability levels. Samples of 500*

Normal probability	Normal probability points		Power for frequency (6·9) with $\lambda_4 = \frac{1}{2}$	
	$c=4$ (upper)	$c=1$ (lower)	$c=4$	$c=1$
(1)	(2)	(3)	(4)	(5)
0·005	3·7062	0·773167	0·1934	0·2067
0·01	3·6094	0·775684	0·2920	0·2790
0·05	3·3766	0·782482	0·5955	0·5196
0·10	3·2695	0·786058	0·7392	0·6509

7. Conclusion and summary

In § 2 of the present paper it is shown that the actual probability of differences between means and variances derived from random samples on the nul-hypothesis may differ considerably from the probability derived from the standard tables (compiled on the assumption that the universal distribution is normal), when, in fact, the universal distribution is *not* normal. Accordingly, the standard tables cannot validly be used unless tests, based on the sample from which the inferences are to be drawn, or on a series of samples produced under similar conditions, have established the likelihood that the universal distribution is approximately normal. In certain cases—but these must be few—the nature of the material may, of itself, suffice to justify the assumption of universal normality. When universal normality cannot be assumed, the best course will be to correct the standard tables using, for this purpose, the moments (up to, say, the fourth) derived from the sample, in conjunction with the formulae given in § 2. This procedure is, of course, open to the objection that the moments derived from the sample may, in fact, differ substantially from the (in general unknown) universal moments, so that any probabilistic inference derived using sample moments must be accepted with reserve. If $b_2 = 3{\cdot}5$, say, it would be safer to assume that the universal value β_2 is 3·5, than to hope (without other evidence) that it is 3, the normal value; it might be 3·75 or even 4, when, usually, the standard table probabilities will be still further astray. It should not be difficult to construct supplementary tables giving very approximate corrections of the standard tables, using the moment expansions given in § 2, for different values of $\sqrt{\beta_1}$ and β_2. To compute unbiassed estimates of the latter, R. A. Fisher's k statistics (1929) should, of course, be used.

It may be asked if testing for normality and, when necessary, correction for universal non-normality is worth the trouble. To answer this question it is desirable to have regard to the logical position of the statistician, concerned with drawing inferences from samples, whose characteristic approach may be defined as *reductio ad paene absurdum*: if an event is highly improbable it must be regarded for practical purposes as impossible. St Thomas Aquinas's* famous 'certitude of probability' is peculiarly apt as applied to the mental attitude of the statistician, from two quite different viewpoints. The first is that decision, and action based on that decision, for which there is not certainty, but merely probabilistic preference, is absolute. One does not say that one has a preference of 20 to 1 for Fertilizer A

* 'According to the Philosopher, certitude is not to be sought equally in every matter.... Hence the certitude of probability suffices, such as may reach the truth in the greater number of cases, although it fails in the minority' (*Summa* IIa–IIae q. lxx, *a.* 2).

over Fertilizer B because the differences between the yields is at or near the 5 % probability point of some test functions: one necessarily decides without qualification that A is better than B.

The second aspect, which has the greater relevance in the present case, is that the statistician regards himself as endowed with 'certitude' when he knows that if he repeated an experiment, as to, say, significant differences in averages, a great number of times, he would be in error in attributing significant difference when, in fact, there was none, in a predetermined proportion of cases. He has certitude as to the probability though his decision in the individual case may be wrong. What is curious is that decisions (which, in effect, are absolute) can be based on probability levels which vary with the temperament of the statistician from perhaps a conservative 0·001 to a daring 0·1. For the particular statistician the probability level will vary with the case: for instance, the present writer would be inclined to suspect non-normality near the 10 % probability level of the $a(1)$ table, whereas he would not be disposed to attach significance in, say, analysis of variance, until about the $2\frac{1}{2}$ % level. Naturally the level will depend on the importance attaching to the decision.

Since all the statistician usually requires from the table of probability for a given measure of significance is whether, on the nul-hypothesis, the probability is 'small', absolute precision is not necessary in the probability. If the probability is thought to be minute, say 0·001, it does not matter if in actual fact it is 0·002 or 0·0005. If, on the contrary, the standard table value is approaching the statistician's level of decision it surely matters a great deal: if he thinks his judgment is likely to be erroneous in 1 out of 20 experiments it must be of importance if, in fact, the true probability is something like 1 in 10 or 1 in 5. These are the kinds of contrasts that appear from § 2, from comparison of standard table probabilities with 'actual' probabilities found when the samples were assumed to be randomly drawn from certain arbitrarily selected types of non-normal universes. The computed probabilities in § 2 admittedly make no claim to exactitude in most of the cases, since the formulae were strained by their application to small sample theory. The point is, however, that the estimates of the actual probabilities are unbiassed in regard to the 'normal theory' probabilities: if the former could be closer to the latter, they might also be further away.

There is one case which is in a quite exceptional category, namely that considered at the beginning of § 2. As far as the writer is aware, this case has never been examined *theoretically* before, despite the extreme simplicity of the algebra. It is shown that in the simplest case of analysis of variance, when the two sample numbers are of the same order of magnitude, the variance is proportional, approximately, to $(\beta_2 - 1)$, so that quite a small measure of universal kurtosis materially changes the probability. Statisticians must have been affected by a kind of hypnosis in favour of normal theory to have overlooked so trivial a point, a stricture from which the writer is not particularly concerned to exclude himself! An exception was E. S. Pearson (1931) who, on the basis of his results cited in § 2 (a), sounded a warning: 'The illustration should serve to emphasize the fact that certain of the "normal theory" tests can be used with greater confidence than others when dealing with samples from populations whose distribution laws are not known.'

An interesting chapter could be written on the fluctuations in the attitude of statisticians during the past century on the question of the occurrence of the normal frequency distribution in nature, a chapter, perhaps, in a large work on Fashions in the Sciences down the Ages. Amongst the following the historian may find the reasons for the prejudice in favour of the hypothesis of universal normality up to, say, the end of the last century:

(1) The fact that, to a close approximation, it applies in a wide range of *mathematical* conditions.

(2) The fact that the theory found practical applications predominantly in assessing the probability of errors in astronomical measurements and in games of chance where the mathematical model could reasonably be assumed to apply.

(3) The beauty of the mathematical theory and the facility of algebraic manipulation in the function involved.

(4) The general shape to the visual sense of such frequency distributions as were known, before χ^2 imposed its discipline.

With the development, about the beginning of the century, of the theory of moments, statisticians became almost over-conscious of universal non-normality. The concomitant semi-invariant approach had quite a different background. The difference between the moment and Karl Pearson curve system on the one hand and semi-invariants and the Gram-Charlier system on the other is fundamentally that for the former normality is a particular case like any other, whereas for the latter normality is basic and generative. Each system has its advantages and disadvantages as applied to the determination of frequency distributions of which the lower moments are known. In fanciful terms one might say that in the ship Gram-Charlier one might sail in perfect safety but only within limited, and more or less ascertainable, range of Port Normality, whereas in the good craft Pearson one can sail the seven seas—at one's own risk.*

Our historian will find a significant change of attitude about a quarter-century ago following on the brilliant work of R. A. Fisher who showed that, when universal normality could be assumed, inferences of the widest practical usefulness could be drawn from samples of any size. Prejudice in favour of normality returned in full force and interest in non-normality receded to the background (though one of the finest contributions to non-normal theory was made during the period by R. A. Fisher himself), and the importance of the underlying assumptions was almost forgotten. Even the few workers in the field (amongst them the present writer) seemed concerned to show that 'universal non-normality doesn't matter': we so wanted to find the theory as good as it was beautiful. References (when there were any at all) in the text-books to the basic assumptions were perfunctory in the extreme. Amends might be made in the interest of the new generation of students by printing in leaded type in future editions of existing text-books and in all new text-books:

Normality is a myth; there never was, and never will be, a normal distribution.

This is an over-statement from the practical point of view, but it represents a safer initial mental attitude than any in fashion during the past two decades.

As already indicated, the present work is incomplete, especially on the experimental side. The writer hopes that he has created a *prima facie* case for the importance of testing for normality.

SUMMARY

(i) Inferences drawn from the standard (normal) tables of z and t may be seriously in error if the conditions in which the standard tables apply (the principal of which is that the universes from which the samples are drawn are normal) are ignored.

* This comment must not be taken as applying to the problem of curve-fitting, i.e. to fitting a smooth curve to given frequencies, but to the problem of estimating the frequency function given the first few semi-invariants.

(ii) Sufficient conditions are given for the approach to normality, with increasing sample size, of the field of tests of normality $a(c)$ (given by (3·1)) for $c > 0$.

(iii) Many term expansions of the first four moments of $a(c)$ for normal samples are given with practical applications designed to find the values of c for which the moments could be used with confidence to find the frequency distributions for medium-size samples; semi-invariants of $a_1(2\cdot 4)$ and $a_1(4)$ ($a_1(c)$ is given by (3·2)) are compared; correlations between $a_1(c)$ and $a_1(c')$ are examined.

(iv) For indefinitely large samples and a wide field of alternative universes $a(4)$ is found to be the most sensitive test of kurtosis and an analogous test of asymmetry $g(c)$ is found to be most sensitive for $c = 3$, $g(3)$ being the familiar $\sqrt{b_1}$.

(v) An examination of the relative efficiency of $a(1)$ and $a(4)$ from the Power Function point of view suggests that $a(4)$ is increasingly to be preferred as the sample size increases; for samples of moderate size $a(1)$ is probably as efficient as $a(4)$.

(vi) Throughout the paper a considerable range of formulae is given in case students may feel interested to carry the writer's researches a stage further so as to give a firmer basis to his conclusions or to modify them. It is suggested (§ 4) that the preparation of a table of probability points of $a(2\cdot 2)$ for normal samples of different sizes be taken in hand.

REFERENCES

Baker, G. A. (1932). *Ann. Math. Statist.* **3**, 1.
Bartlett, M. S. (1935). *Proc. Camb. Phil. Soc.* **31**, 226.
Cornish, E. A. & Fisher, R. A. (1937). *Rev. Inst. Int. Statist.* **5**, 307.
Craig, C. C. (1928). *Metron*, **7**, 3.
Eden T. & Yates, F. (1933). *J. Agric. Sci.* **23**, 6.
Fisher, R. A. (1925). *Metron*, **5**, 90.
Fisher, R. A. (1929). *Proc. Lond. Math. Soc.* (2), **30**, 199.
Fisher, R. A. (1930). *Proc. Roy. Soc.* A, **130**, 16.
Fisher, R. A. & Wishart, J. (1931). *Proc. Lond. Math. Soc.* (2), **33**, 195.
Fréchet, M. (1937). *Géneralités sur les Probabilités. Variables aléatoires.*
Geary, R. C. (1933). *Biometrika*, **25**, 184.
Geary, R. C. (1935). *Biometrika*, **27**, 310, 353.
Geary, R. C. (1936). *J. Roy. Statist. Soc.* (*Supplement*), **3**, 178.
Geary, R. C. (1936). *Biometrika*, **28**, 295.
Geary, R. C. (1947). *Biometrika*, **34**, 68.
Geary, R. C. & Pearson, E. S. (1938). *Tests of Normality.*
Geary, R. C. & Worlledge, J. P. G. (1946). *Biometrika*, **34**, 98.
Gosset, W. S. (1908). *Biometrika*, **6**, 1.
Hsu, C. T. & Lawley, D. N. (1940). *Biometrika*, **31**, 238.
Kendall, M. G. (1941). *Biometrika*, **32**, 81.
Nair, A. N. K. (1942). *Sankhyā*, **5**, 393.
Neyman, J. & Pearson, E. S. (1933). *Philos. Trans.* A, **231**, 289.
Neyman, J. & Pearson, E. S. (1936). *Statist. Res. Mem.* **1**, 1.
Pearson, E. S. (1929). *Biometrika*, **21**, 337.
Pearson, E. S. (1930). *Biometrika*, **22**, 239.
Pearson, E. S. (1931*a*). *Biometrika*, **22**, 423.
Pearson, E. S. (1931*b*). *Biometrika*, **23**, 114.
Pearson, E. S. (1935). *Biometrika*, **27**, 333.
Pearson, E. S. & Adyanthaya, N. K. (1929). *Biometrika*, **21**, 259.
Pearson, Karl (1895). *Philos. Trans.* A, **186**, 343.
Pepper, Joseph (1932). *Biometrika*, **24**, 55.
Rider, P. R. (1931). *Ann. Math. Statist.* **2**, 48.
Rietz, H. L. (1939). *Ann. Math. Statist.* **10**, 265.
Shewhart, W. A. & Winters, F. W. (1928). *J. Amer. Statist. Ass.* **23**, 144.
Wishart, J. (1930). *Biometrika*, **22**, 224.
Yasukawa, K. (1934). *Tokohu Math. J.* **38**, 465.

THE STRATIFIED SEMI-STATIONARY POPULATION

By S. VAJDA

1. Constant population

Let a set of non-increasing real values $p_0 = 1,\ p_1, \ldots, p_n, p_{n+1} = 0$ be given, and let p_i represent the probability of a person of age 0 surviving the i following years. Further, let $l_0, l_1, \ldots, l_n$ represent the numbers of persons of age $0, 1, \ldots, n$ living at time $t = 0$. We consider then the development of such a population during the years following $t = 0$, under the assumption that the probabilities p_i remain the same throughout the period investigated.

Only persons of age 0 are to enter the population, and the number of such entrants shall be such that the total of the population is kept constant at a number $H = \sum_{i=0}^{n} l_i$. At the end of the first year the survivors of the H persons who were alive at $t = 0$ will be (if we put $l_i/p_i = r_i$, say)

$$\sum_{i=0}^{n-1} l_i \frac{p_{i+1}}{p_i} = \sum_{i=0}^{n-1} r_i p_{i+1} \leqslant H,$$

and therefore the number of entrants at the beginning of the second year (i.e. at $t = 1$) is $\phi_1 = H - \sum_{i=0}^{n-1} r_i p_{i+1}$.

By the same argument the entrants at $t = 2$ will be

$$\phi_2 = H - \phi_1 p_1 - \sum_{i=0}^{n-2} r_i p_{i+2},$$

and so on; generally

$$\phi_t = H - \phi_{t-1} p_1 - \phi_{t-2} p_2 - \ldots - \phi_1 p_{t-1} - \sum_{i=0}^{n-t} r_i p_{i+t}, \tag{1}$$

as long as $t \leqslant n$, that is, as long as there are survivors of the initial population. For $t > n$ we obtain

$$H = \phi_t + \phi_{t-1} p_1 + \phi_{t-2} p_2 + \ldots + \phi_{t-n} p_n. \tag{2}$$

We want to find an expression for ϕ_t, which must obviously depend on $l_1, l_2, \ldots, l_n$. Now (2) is a difference equation for the function ϕ_t of t and can easily be solved. For this purpose consider the 'characteristic equation'

$$x^n + x^{n-1} p_1 + x^{n-2} p_2 + \ldots + x p_{n-1} + p_n = 0. \tag{3}$$

Let this equation have the roots $x_1, x_2, \ldots, x_r$, where x_i is a k_i-fold root and $x_i \neq x_j$. We have then as a solution of the difference equation (2)

$$\phi_t = H_1 + P_1(t)\, x_1^t + \ldots + P_r(t)\, x_r^t, \tag{4}$$

where $H_1 = H/\Sigma p_i$ and $P_i(t) = \alpha_{i1} + \alpha_{i2} t + \ldots + \alpha_{ik_i} t^{k_i - 1}$. The α_{ij} must be found from the initial population, i.e. from equations of the form (1) which contain the first n numbers of entrants $\phi_1, \phi_2, \ldots, \phi_n$. But we find by inspecting these equations, which are of the form ($t \leqslant n$)

$$H = \phi_t + \phi_{t-1} p_1 + \ldots + \phi_1 p_{t-1} + r_0 p_t + r_1 p_{t+1} + \ldots + r_{n-t} p_n,$$

that they are equivalent to

$$r_t = \phi_{-t} = H_1 + \sum_{i=1}^{r} \sum_{j=1}^{k_i} \alpha_{ij} x_i^{-t}(-t)^{j-1}. \tag{5}$$

Hence the α_{ij} can be fixed, dependent on the $r_i = l_i/p_i$ and thus on the initial population.

We have thus proved:

If a population with an age distribution $l_0, l_1, \ldots, l_n$ is subject to survival rates p_i ($i = 1, 2, \ldots, n$), and if this population is kept constant by ϕ_t entrants of age 0 at the end of the tth year, then ϕ_t is given by (4), where the x_i are the different roots of (3), and the α_{ij} must be found from the set (5).

The population after t years will have the following age distribution:

$$\phi_t, \quad \phi_{t-1}p_1, \quad \phi_{t-2}p_2, \quad \ldots, \quad \phi_{t-n}p_n.$$

It can easily be proved that, if the p_i are decreasing (and not merely non-increasing), then for all the roots x_i of equation (3) we have $|x_i| < 1$ and that any real root must be negative. Hence the ϕ_t will oscillate around their limit $\lim_{t=\infty} \phi_t = H_1$. The age distribution of the population thus tends, again through oscillations, to $H_1, H_1 p_1, \ldots, H_1 p_n$, which may be called the *intrinsic stationary population*. Obviously, if the initial population has already this distribution, it will not alter any more and the number of entrants will be constant and $= H_1$. In such a case all $\alpha_{ij} = 0$, and $r_i = H_1 = l_0$, whatever the x_i may be.

On the other hand, if $p_i = p_{i+1}$ holds for one or more values of i, then we may get cycles, and this is easily seen for the equation $x^n + x^{n-1} + \ldots + x + 1 = 0$. All roots have modulus 1, and it depends on the initial population whether we are dealing with the stationary case or with periodic cycles. No tendency towards an intrinsic stationary population appears in such a case.

Example. Let us assume that we have the following probabilities of survival:

p_1	p_2	p_3	p_4	p_5	p_6
7/8	49/96	5/32	13/384	1/384	0

The characteristic equation can then be written

$$384x^5 + 336x^4 + 196x^3 + 60x^2 + 13x + 1 = 0,$$

which has the five different roots

$$-\tfrac{1}{4} \pm \sqrt{-\tfrac{5}{48}}, \quad -\tfrac{1}{8} \pm \sqrt{-\tfrac{7}{64}} \quad \text{and} \quad -\tfrac{1}{8}.$$

The initial population will be assumed to be

l_0	l_1	l_2	l_3	l_4	l_5
859	1269	229	50	115	56

which implies $H_1 = 1000$ (approx.). Therefore the $r_i = l_i/p_i$ are

r_0	r_1	r_2	r_3	r_4	r_5
859	1450·38	448·86	319·14	3405·42	21420·90

From $r_k = 1000 + \alpha_1 x_1^{-k} + \alpha_2 x_2^{-k} + \ldots + \alpha_5 x_5^{-k}$ ($k = 1, 2, \ldots, 5$) we find

α_1	α_2	α_3	α_4	α_5
$-70+60i$	$-70-60i$	0	0	-1

It follows that the number of entrants in the year t ($= 0, 1, 2, \ldots$) will be

$$\phi_t = 1000\left[1 + (-0{\cdot}07 + 0{\cdot}06i)\left(-\tfrac{1}{4} + \sqrt{-\tfrac{5}{48}}\right)^t + (-0{\cdot}07 - 0{\cdot}06i)\left(-\tfrac{1}{4} - \sqrt{-\tfrac{5}{48}}\right) + \frac{-0{\cdot}001}{(-8)^t}\right].$$

These numbers are given in the first row of Table 1, which shows the evolution of the whole population.

Table 1

t	0	1	2	3	4	5	6	7	8 and after
Age 0	859	996	1025	989	1001	1001	999	1000	1000
1	1269	752	872	897	865	876	876	874	875 (= 1000 × 7/8)
2	229	740	438	508	523	505	511	511	510 (= 1000 × 49/96)
3	50	70	227	134	156	160	155	156	156 (= 1000 × 5/32)
4	115	11	15	49	29	34	34	34	34 (= 1000 × 13/384)
5	56	9	1	1	4	2	3	3	3 (= 1000 × 1/384)
	2578	2578	2578	2578	2578	2578	2578	2578	2578

2. Two constant populations

All this covers well-known ground.* A new problem arises, however, when we consider two initial populations with two sets of probabilities of survival, say p_i ($i = 1, 2, \ldots, n_1$) and $\bar{p}_i$ ($i = 1, 2, \ldots, n_2$), where $p_0 = \bar{p}_0 = 1$ and $p_{n_1} . p_{n_2} \neq 0$. We ask now whether it is possible to keep both constant by the same number of yearly entrants. More precisely:

Let the two equations

$$\sum_{i=0}^{n_1} p_i x^{n_1 - i} = 0 \quad \text{and} \quad \sum_{i=0}^{n_2} \bar{p}_i y^{n_2 - 1} = 0$$

have the roots $x_1, \ldots, x_r$ with multiplicities $k_1, \ldots, k_r$ and $y_1, \ldots, y_s$ with multiplicities $j_1, \ldots, j_s$ respectively. No two x_i or two y_i are equal and no x_i or y_i is zero. Under what further conditions, concerning the x's and the y's, can the expressions ϕ_t and ψ_t then have the same numerical values for all integral values of t, i.e.

$$(H_1 - \bar{H}_1) + \sum_{i=1}^{r} P_i(t)\, x_i^t - \sum_{i=1}^{s} \bar{P}_i(t)\, y_i^t = 0 \quad \text{for} \quad t = 0, 1, 2, \ldots, \tag{6}$$

where $$\phi_t = H_1 + \sum_{i=1}^{r} P_i(t)\, x_i^t \quad \text{with} \quad P_i(t) = \alpha_{i1} + \alpha_{i2} t + \ldots + \alpha_{ik_i} t^{k_i - 1}$$

and $$\psi_t = \bar{H}_1 + \sum_{i=1}^{s} \bar{P}_i(t)\, y_i^t \quad \text{with} \quad \bar{P}_i(t) = \beta_{i1} + \beta_{i2} t + \ldots + \beta_{ij_i} t^{j_i - 1}?$$

Suppose first that none of the x's equals any of the y's. Then it is known that the determinant of any set of equations of the system (6) is not zero. It follows that we must have $H_1 = \bar{H}_1$ and all α's and β's $= 0$, hence all $P_i(t)$ and $\bar{P}_i(t) \equiv 0$. In this case the two populations must already be stationary and therefore identical with the intrinsic stationary populations which are implied by the sets p_i and $\bar{p}_i$, respectively.

On the other hand, if some of the x's are equal to some of the y's, say $x_1 = y_1, \ldots, x_m = y_m$ and all the others are different, then we find by the same argument that $H_1 = \bar{H}_1$ and $P_i = \bar{P}_i$ for the first m values of i, whereas all the other P_i and $\bar{P}_i$ are identically zero. (It is, of course, again possible that *all* the P_i and $\bar{P}_i$ are identically zero and that we have, in fact, again the two intrinsic stationary populations.)

If all x's are equal to the y's, with equal multiplicities, then the two equations are equal

* It follows, for example, from results of P. H. Leslie (1945).

and the two populations must be identical, if they are to be kept constant by equal numbers of entrants.

We have thus reached the following conclusion: If we assume that the two equations given above are not identical, and that the initial populations are not the intrinsic stationary ones, then they must be such that the two equations have some (but not all) roots equal and if we calculate the corresponding P_i and $\bar{P}_i$ (see (4) and (5) of the previous section), then those corresponding to the equal roots must be identical and the others must vanish. This includes the case where the x's and y's are the same, but with different multiplicities, so that the P_i and $\bar{P}_i$ do not all extend to the highest power of t which would be admissible by (7) or (8) respectively.

If the two populations are the intrinsic stationary ones, then the numbers of entrants will be constant (i.e. independent of the year) and the two constants will be equal if and only if

$$\frac{\Sigma l_i}{\Sigma p_i} = \frac{\Sigma \bar{l}_i}{\Sigma \bar{p}_i},$$

where the first expression refers to the first and the second expression to the second population.

Example. In the example used in § 1 we have $\alpha_3 = \alpha_4 = 0$, and we can therefore try to obtain a second population which is kept constant by the same numbers of entrants as the first one. We construct an equation which has again the roots $-\frac{1}{4} \pm \sqrt{-\frac{5}{48}}$ and $-\frac{1}{8}$, but not $-\frac{1}{8} \pm \sqrt{-\frac{7}{64}}$. Such an equation is, for example,

$$480x^5 + 396x^4 + 218x^3 + 62x^2 + 13x + 1 = 0,$$

which has the roots $-\frac{1}{4} \pm \sqrt{-\frac{5}{48}}$, $-\frac{1}{8}$, and $-\frac{1}{10} \pm \sqrt{-\frac{9}{100}}$. It implies that the probabilities of survival, i.e. $\bar{p}_i$, are

$$33/40 \quad 109/240 \quad 31/240 \quad 13/480 \quad 1/480,$$

and as the ϕ_t (and the $r_t = \phi_{-t}$) are to be the same as in § 1, the initial population must now be $\bar{l}_i = r_i \bar{p}_i$. Table 2 shows the development of such a population, and it will be seen that the first line is identical with that in Table 1.

Table 2

t	0	1	2	3	4	5	6	7	8 and after
Age 0	859	996	1025	989	1001	1001	999	1000	1000
1	1196	709	822	845	816	826	826	824	825 (= 1000 × 33/40)
2	204	658	390	452	465	449	455	455	454 (= 1000 × 109/240)
3	41	58	187	111	129	132	128	129	129 (= 1000 × 31/240)
4	92	9	12	39	23	27	27	27	27 (= 1000 × 13/480)
5	45	7	1	1	3	2	2	2	2 (= 1000 × 1/480)
	2437	2437	2437	2437	2437	2437	2437	2437	2437

3. STRATIFIED POPULATION: TWO GRADES

The results of the previous sections will now be used for an investigation of the stratified population.* First, we consider a population split into a lower and a higher grade in the following way:

We assume that all members of age 0 are in the lower grade only, but that all other ages may share in both grades. Apart from mortality, which operates on all members according to their age, we assume that at every age a certain proportion dependent on that age is 'promoted', at the end of the year, from the lower into the higher grade. Our problem is to discover whether this can be done whilst maintaining the totals in both grades constant; naturally the grand total of the population must remain constant.

It is sufficient to deal only with the lower grade, as the numbers at each age in the higher one can be found by subtracting those in the lower grade from the total population at that age. Now the lower grade is depleted by mortality and also by promotions. If the probability of remaining unpromoted until age i is t_i, then the probability of not leaving the grade in this period is $p_i t_i = \bar{p}_i$, say. Since all entrants into the population are at the same time entrants into the lower grade, our problem thus reduces to the following:

Is it possible to find an initial population, stratified into two grades, such that, on the basis of mortality described by p_i, the number of entrants every year necessary to keep the population constant is the same as that calculated on the basis of mortality-cum-promotion, described by $\bar{p}_i$?

We can apply our results in § 2 to this case by considering the lower grade and the total population as the two populations given. It follows that the lower grade can only be kept constant by that number of entrants which is necessary for the total population, if the latter is initially such that some of the $P_i(t)$ which depend on it are either identically zero or at least do not extend to the highest degree indicated by the multiplicities of the corresponding roots in $\Sigma p_i x^{n-i} = 0$. In order to find a suitable initial population for the lower grade it is then necessary to find an equation $\Sigma \bar{p}_i y^{n-i} = 0$ which has the roots, with the necessary multiplicities, which appear explicitly in ϕ_t as calculated from the original equation, but which is not identical with it. The degree of $\Sigma \bar{p}_i y^{n-i} = 0$ may be lower than or equal to that of $\Sigma p_i x^{n-i} = 0$. If it is lower, then all members of the population will be in the higher grade at the highest age or ages.

This condition is not sufficient, however. In view of the interpretation of the equation containing the $\bar{p}_i$'s these coefficients must be positive and, as the lower grade is a part of the whole, we must have $\bar{p}_i \leqslant p_i$ for all i. But it is not necessary that we have also $\bar{p}_{i+1} \leqslant \bar{p}_i$. If the opposite holds, this could still bear a practical interpretation. It would mean that reversions occur from the higher into the lower grade.

If an equation with the necessary and sufficient properties can be found, then we take the $r_i = l_i/p_i$ which we had to start with and construct the initial population of the lower grade by writing the number at age i as $\bar{l}_i = r_i \bar{p}_i = l_i \bar{p}_i/p_i$.

It will be seen that in such a population the age distributions change with the passage of time (tending to a stationary limit) but that nevertheless all entrants have the same combined prospects of survival and promotion. (Thus from the point of view of a member of the community his position is the same as if he entered a stationary population. His chances

* Cf., for the stationary case, with continuous changes, H. L. Seal (1945).

of promotion are unaffected by the changes in the age distribution of those in front of him. But the characteristics of the population as a whole, for instance the efficiency of the staff from the point of view of an employer may, of course, vary considerably.) Such a population will be called *semi-stationary*.

Example. The population shown in Table 2 can be taken as representing a lower grade within the population given in Table 1. The ratios $t_i = \bar{p}_i/p_i$ are then:

t_0	t_1	t_2	t_3	t_4	t_5
1	33/35	218/245	62/75	4/5	4/5

Table 3 is constructed by subtracting Table 2 from Table 1 and thus shows the composition of the higher grade.

Table 3

t	0	1	2	3	4	5	6	7	8 and aft
Age 1	73	43	50	52	49	50	50	50	50
2	25	82	48	56	58	56	56	56	56
3	9	12	40	23	27	28	27	27	27
4	23	2	3	10	6	7	7	7	7
5	11	2	—	—	1	—	1	1	1
	141	141	141	141	141	141	141	141	141

4. Stratified population: more than two grades

Let us now split up the higher grade as well. We have then, say, k grades, with grades 2 and above forming the aggregate which was simply called the higher grade in § 2; grade 1 is identical with the lower grade of that section.

We assume further that promotions from any grade into the next higher one take place at the end of every year and that every promotee into any grade has to stay there for at least one year. Thus in any population the lowest possible age of grade g is $g-1$. The actual lowest ages may be different, because the first promotion rates different from 0 may concern higher ages than these. The rates of promotion can be different from grade to grade, but depend within each grade only on the age, as before.

We shall again investigate whether it is possible to keep the total numbers of every grade constant, even if the age distributions of the grades are changing.

We have seen that the age distribution of the total population, after t years, is

$$\phi_t, \quad \phi_{t-1}p_1, \quad \ldots, \quad \phi_{t-n}p_n.$$

The distribution of grade 1 is, at the same time,

$$\phi_t, \quad \phi_{t-1}\bar{p}_1, \quad \ldots, \quad \phi_{t-n}\bar{p}_n,$$

and it is assumed that the set of p_i is not identical with the set of $\bar{p}_i$. Hence grades 2 and above will have the age distribution

$$\phi_t(p_0-\bar{p}_0), \quad \phi_{t-1}(p_1-\bar{p}_1), \quad \phi_{t-n}(p_n-\bar{p}_n).$$

Let us assume that $\phi_{t-\nu}(p_\nu-\bar{p}_\nu)$ is the first item in this series which is not zero. Clearly we have $\nu \geqslant 1$. Then, as far as numbers of members (and not their individual careers) are con-

cerned, this aggregate of grades 2 and above is equivalent to a population which has arisen from successive annual entrants $\phi_{t-\nu}(p_\nu - \bar{p}_\nu)$ who have been subject to rates of survival

$$q_1 = \frac{p_{\nu+1} - \bar{p}_{\nu+1}}{p_\nu - \bar{p}_\nu}, \quad \ldots, \quad q_{n-\nu} = \frac{p_n - \bar{p}_n}{p_\nu - \bar{p}_\nu}.$$

It must be understood, however, that 'survival' is here a balance between deaths and promotions into the grade, so that these rates may very well exceed unity.

The number of annual entrants into grade 2 is given by

$$\phi_{t-\nu}(p_\nu - \bar{p}_\nu) = \left[H_1 + \sum_{i=1}^{m} P_i(t-\nu)\, x_i^{t-\nu}\right](p_\nu - \bar{p}_\nu),$$

where $x_1, \ldots, x_m$ are the common roots of $\sum_{i=0}^{n} p_i x^{n-i} = 0$ and $\sum_{i=1}^{n} \bar{p}_i y^{n-i} = 0$, with multiplicities k_i and j_i respectively, and where the P_1 are polynomials whose order does not exceed either $k_i - 1$ or $j_i - 1$. (They may all be identically zero.)

The x_i are, of course, also roots of $\sum_{i=\nu}^{n} (p_i - \bar{p}_i)\, x^{n-i} = 0$, with multiplicities given by the smaller of k_i and j_i.

We ask now if it is possible to construct grade 2 alone in such a way that its total remains also constant. The argument which has been used in § 3 shows that this is possible if another equation of grade $n - \nu$ can be found whose coefficients w_i, say, are not larger than the corresponding q_i (and $w_0 = 1$), which has once again the roots $x_1, \ldots, x_m$, with multiplicities g_i at least. If $g_1 + \ldots + g_m = n - \nu$, then this is clearly impossible. If $g_1 + \ldots + g_m$ is smaller than this value, then we can try to find such an equation. The initial population can also be then found, if we multiply the initial population of grades 2 and above by w_i/q_i. The grades 1, 2 and the aggregate of 3 and above can then be constructed and every stratum kept constant, but with changing age distributions.

We can proceed in the same way and find at each step whether further splitting up is possible beyond 3 grades, 4 grades, etc. It is seen that in general, if $\Sigma g_i = n - m$, and if grade g starts in fact at age $g - 1$, then $m + 1$ grades can exist.

The smallest value of Σg_i is 1, and in this extreme case n grades can be constructed, i.e. one less than the number of ages. The nth grade will then contain the ages $n - 1$ and n. Further, since x_1 is a root of $x - x_1 = 0$, the age distribution of this highest grade is

$$H_1 + \alpha_1 x_1^{t-n+1}, \quad -(H_1 + \alpha_1 x_1^{t-n})\, x_1 = -H_1 x_1 - \alpha_1 x_1^{t-n+1}$$

(x_1 is, of course, negative).

Example. We use again the same example as before. The characteristic equation for the whole population was

$$x^5 + \tfrac{7}{8}x^4 + \tfrac{49}{96}x^3 + \tfrac{5}{32}x^2 + \tfrac{13}{384}x + \tfrac{1}{384} = 0,$$

and that for grade 1 alone

$$x^5 + \tfrac{33}{40}x^4 + \tfrac{109}{240}x^3 + \tfrac{31}{240}x^2 + \tfrac{13}{480}x + \tfrac{1}{480} = 0.$$

The difference between these two equations gives the equation for grade 2 and above

$$x^4 + \tfrac{9}{8}x^3 + \tfrac{13}{24}x^2 + \tfrac{13}{96}x + \tfrac{1}{96} = 0.$$

This equation has, of course, the roots $-\frac{1}{4} \pm \sqrt{-\frac{5}{48}}$, $-\frac{1}{8}$ which are common to the two characteristic equations of the fifth degree, and also a further root $-\frac{1}{2}$. Now there is

a biquadratic equation with the three specified common roots and not larger coefficients and having the coefficient of x^4 equal to unity), viz.

$$x^4 + \tfrac{33}{40}x^3 + \tfrac{17}{48}x^2 + \tfrac{1}{15}x + \tfrac{1}{240} = 0.$$

The fourth, irrelevant, root is $-1/5$. This equation leads to the following development:

Grade 2 only

t	0	1	2	3	4	5	6	7	8 and after
Age 1	73	43	50	52	49	50	50	50	50
2	18	60	35	41	43	40	41	41	41
3	6	8	26	14	18	19	18	18	18
4	11	—	1	5	2	3	3	3	3
5	4	1	—	—	—	—	—	—	—
	112	112	112	112	112	112	112	112	112

Grade 3 and above

Age 2	7	22	13	15	15	16	15	15	15
3	3	4	14	9	9	9	9	9	9
4	12	2	2	5	4	4	4	4	4
5	7	1	—	—	1	—	1	1	1
	29	29	29	29	29	29	29	29	29

Analysis into further grades is impossible in this case, because the characteristic equation of the third grade does not have any roots apart from the three common roots of all previous equations.

5. Promotion rates dependent on seniority

We still consider more than two grades, but now we will assume that the promotion rates do not depend on the attained age but on the seniority, i.e. on the time spent in the grade, instead. In the lowest grade seniority is equivalent to age, because all members were supposed to enter at the lowest age only. If we consider again the two grades of § 3, but this time take note of differences in seniority, we find the following pattern:

Age	Lower grade	Higher grade: Seniority 0	1	...	$x-1$	Total
0	ϕ_t					ϕ_t
1	$\phi_{t-1}p_1t_1$	$\phi_{t-1}p_1(t_0-t_1)$				$\phi_{t-1}p_1$
2	$\phi_{t-2}p_2t_2$	$\phi_{t-2}p_2(t_1-t_2)$	$\phi_{t-2}p_2(t_0-t_1)$			$\phi_{t-2}p_2$
...	....	....	....	...	....	....
x	$\phi_{t-x}p_xt_x$	$\phi_{t-x}p_x(t_{x-1}-t_x)$	$\phi_{t-x}p_x(t_{x-2}-t_{x-1})$	...	$\phi_{t-x}p_x(t_0-t_1)$	$\phi_{t-x}p_x$
...	....	....	....	...	....	....

Note. $t_i = \bar{p}_i/p_i$ and hence $t_0 = 1$.

If we consider now promotion from grade 2 into grade 3, and if we introduce u_s, the probability of not being promoted during s years from grade 2 ($u_0 = 1$), we see that grades 2 and 3 (including higher grades, if any) will have the following constitution:

Grade 2

Age	Seniority 0	1		$x-1$
1	$\phi_{t-1}p_1(t_0-t_1)\,u_0$			
2	$\phi_{t-2}p_2(t_1-t_2)\,u_0$	$\phi_{t-2}p_2(t_0-t_1)\,u_1$		
...	....	....	...	....
x	$\phi_{t-x}p_x(t_{x-1}-t_x)\,u_0$	$\phi_{t-x}p_x(t_{x-2}-t_{x-1})\,u_1$	...	$\phi_{t-x}p_x(t_0-t_1)\,u_{x-1}$
...	....	....	...	....

Grade 3

Age	Seniority 0	1		$x-2$
2	$\phi_{t-2}p_2(t_0-t_1)\,(u_0-u_1)$			
3	$\phi_{t-3}p_3[(t_1-t_2)\,(u_0-u_1)$ $+(t_0-t_1)\,(u_1-u_2)]$	$\phi_{t-3}p_3(t_0-t_1)\,(u_0-u_1)$		
...	....	....	...	....
x	$\phi_{t-x}p_x[(t_{x-2}-t_{x-1})\,(u_0-u_1)$ $+(t_{x-3}-t_{x-2})\,(u_1-u_2)+\dots$ $+(t_0-t_1)\,(u_{x-2}-u_{x-1})]$	$\phi_{t-x}p_x[(t_{x-3}-t_{x-2})\,(u_0-u_1)$ $+(t_0-t_1)\,(u_{x-3}-u_{x-2})]$	...	$\phi_{t-x}p_x(t_0-t_1)\,(u_0-u_1)$
...	....	....	...	....

It follows by means of the same argument as before that grade 2 can be kept constant if we can find the u_i such that the equation

$$x^{n-1}p_1(t_0-t_1)+x^{n-2}p_2[(t_1-t_2)+(t_0-t_1)\,u_1]+\dots$$
$$+p_n[(t_{n-1}-t_n)+(t_{n-2}-t_{n-1})\,u_1+\dots+(t_0-t_1)\,u_{n-1}] = 0$$

has the same roots which were common to $x^n+x^{n-1}p_1+\dots+p_n = 0$ and

$$x^{n-1}p_1(t_0-t_1)+x^{n-2}p_2(t_0-t_2)+\dots+p_n(t_0-t_n) = 0,$$

which is identical with the difference of the first two equations of degree n, referring respectively to the whole population and to the lowest grade. We must further insist that all u_i must have non-negative values, not larger than 1. The coefficients of the powers of x must also be positive, but it is not necessary that $u_{i+1} \leqslant u_i$, unless we do not admit reversions. If m is the number of common roots, then it follows again as in the last section that $n-m+1$ grades could exist which remain constant under the operation of promotions, but that their age and seniority distributions change.

Example. Dealing once more with the same example as in the previous sections, we have to find a biquadratic equation

$$\tfrac{7}{8}(1-\tfrac{33}{35})\,x^4+\tfrac{49}{96}[(\tfrac{33}{35}-\tfrac{218}{245})+(1-\tfrac{33}{35})\,u_1]\,x^3+\tfrac{5}{32}[(\tfrac{218}{245}-\tfrac{62}{75})+(\tfrac{33}{35}-\tfrac{218}{245})\,u_1+(1-\tfrac{33}{35})\,u_2]\,x^2$$
$$+\tfrac{13}{384}[(\tfrac{62}{75}-\tfrac{4}{5})+(\tfrac{218}{245}-\tfrac{62}{75})\,u_1+(\tfrac{33}{35}-\tfrac{218}{245})\,u_2+(1-\tfrac{33}{35})\,u_3]\,x$$
$$+\tfrac{1}{384}[(\tfrac{4}{5}-\tfrac{4}{5})+(\tfrac{62}{75}-\tfrac{4}{5})\,u_1+(\tfrac{218}{245}-\tfrac{62}{75})\,u_2+(\tfrac{33}{35}-\tfrac{218}{245})\,u_3+(1-\tfrac{33}{35})\,u_4] = 0,$$

or, if we use four significant figures in every fraction,

$$\begin{aligned} x^4 &+ (0{\cdot}5417 + 0{\cdot}5833u_1)\,x^3 + (0{\cdot}1973 + 0{\cdot}1658u_1 + 0{\cdot}1786u_2)\,x^2 \\ &+ (0{\cdot}01805 + 0{\cdot}04274u_1 + 0{\cdot}03593u_2 + 0{\cdot}03869u_3)\,x \\ &+ (0 + 0{\cdot}001389u_1 + 0{\cdot}003288u_2 + 0{\cdot}002764u_3 + 0{\cdot}002976u_4) = 0. \end{aligned}$$

This biquadratic equation must have the roots $-\frac{1}{4} \pm \sqrt{-\frac{5}{48}}$ and $-\frac{1}{8}$. If the fourth root is called $(-z)$, then the equation must be identical with

$$(x^3 + \tfrac{5}{8}x^2 + \tfrac{11}{48}x + \tfrac{1}{48})\,(x+z) = 0.$$

Simple arithmetic shows then that

$$u_1 = 0{\cdot}1429 + 1{\cdot}7142z, \quad u_2 = 0{\cdot}0459 + 1{\cdot}9082z,$$

$$u_3 = -0{\cdot}1283 + 2{\cdot}2566z \quad \text{and} \quad u_4 = 0{\cdot}0019 + 1{\cdot}9962z.$$

Now z must be at least $0{\cdot}05685$ to make u_3 positive and it must not exceed $0{\cdot}5$, because otherwise the u_i would exceed unity. But then u_4 will always be larger than u_3, unless we put $z = \frac{1}{2}$ which would mean $u_i = 1$ for all i and then there would be no members at all in grades 3 and above. It follows that we must admit reversions from grade 3 into grade 2. We can then, for instance, take $z = 0{\cdot}2$ and have

$$u_1 = 0{\cdot}4857, \quad u_2 = 0{\cdot}4275, \quad u_3 = 0{\cdot}3230 \quad \text{and finally} \quad u_4 = 0{\cdot}4011.$$

The biquadratic equation becomes

$$x^4 + \tfrac{33}{40}x^3 + \tfrac{17}{48}x^2 + \tfrac{1}{15}x + \tfrac{1}{240} = 0.$$

This is the same as the one used in § 4, and we can again write down the changing pattern of the population, but this time taking also seniority into account:

Grade 2

Age	$t=0$						1					2					3				
1	73	—	—	—	—	73	43	—	—	—	43	50	—	—	—	50	52	—	—	—	52
2	12	6	—	—	—	18	40	20	—	—	60	23	12	—	—	35	27	14	—	—	41
3	3	2	1	—	—	6	4	2	2	—	8	15	6	5	—	26	8	3	3	—	14
4	3	3	3	2	—	11	—	—	—	—	0	—	1	—	—	1	1	2	1	1	5
5	—	—	2	1	1	4	—	—	1		1	—	—	—	—	—	—	—	—	—	—
	91	11	6	3	1	112	87	22	3	-	112	88	19	5	—	112	88	19	4	1	112

Age	4						5					6					7 and later				
1	49	—	—	—	—	49	50	—	—	—	50	50	—	—	—	50	50	—	—	—	50
2	28	15	—	—	—	43	27	13	—	—	40	27	14	—	—	41	27	14	—	—	41
3	10	4	4	—	—	18	10	5	4	—	19	10	4	4	—	18	10	4	4	—	18
4	1	1	—	—	—	2	1	1	1	—	3	1	1	1	—	3	1	1	1	—	3
5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	88	20	4	—	—	112	88	19	5	—	112	88	19	5	—	112	88	19	5	—	112

Grade 3

Age	0						1					2					3				
2	7	—	—	—	—	7	22	—	—	—	22	13	—	—	—	13	15	—	—	—	15
3	1	2	—	—	—	3	2	2	—	—	4	6	8	—	—	14	4	5	—	—	9
4	4	3	5	—	—	12	1	—	1	—	2	—	1	1	—	2	1	2	2	—	5
5	1	2	2	2	—	7	—	—	—	1	1	—	—	—	—	—	—	—	—	—	—
	13	7	7	2	—	29	25	2	1	1	29	19	9	1	—	29	20	7	2	—	29

Age	4						5					6					7 and later				
2	15	—	—	—	—	15	16	—	—	—	16	15	—	—	—	15	15	—	—	—	15
3	4	5	—	—	—	9	4	5	—	—	9	4	5	—	—	9	4	5	—	—	9
4	1	1	2	—	—	4	1	1	2	—	4	1	1	2	—	4	1	1	2	—	4
5	—	1	—	—	—	1	—	—	—	—	—	—	1	—	—	1	—	1	—	—	1
	20	7	2	—	—	29	21	6	2	—	29	20	7	2	—	29	20	7	2	—	29

We find, as before, that further splitting up of grades is impossible, if the total in each grade is to remain constant throughout the years.

Summary

This investigation deals with a stratified population, which is subject to (i) mortality, dependent on age, and to (ii) promotion rates, indicating the ratios of members of a grade which are transferred to the next higher grade at the end of the year.

Section 1 concerns a population which is not yet stratified and formulae are deduced to calculate the number of entrants at time t, necessary to replace yearly deaths and thus to keep the total of the population constant. This number depends clearly on the mortality rates and on the age distribution existing at time $t = 0$. In general the population tends towards a limiting age distribution, the 'intrinsic stationary population'.

Section 2 considers two populations and conditions are derived for the case that they need, every year, equal numbers of entrants to keep them constant.

Section 3 introduces the stratified population. Both mortality and promotion rates depend on the age, and they are independent of the time t. Under certain conditions one of the two populations considered in § 2 can be taken as the whole and the other as the lowest grade in it. It is shown how and when entries into the grade can, at the same time, replace both losses due to mortality in the whole population, and to mortality and promotion depleting the lowest grade. This can also be described by saying that the totals of both grades can be kept constant at the same time, although the age distributions change from year to year.

Section 4 generalizes the results of the previous section for a population consisting of k grades. If the population is spread over n ages, then it is shown that up to $n-1$ grades

may be possible in the most favourable case, such that they are all kept constant, whilst the age distributions all oscillate. Such a population is called semi-stationary.

Section 5 introduces the case which has been of actual importance in practical establishment work: the promotion rates are made dependent on the time spent in the grade instead of on the age.

A numerical example is attached to § 1 and is carried through all stages to illustrate the results which emerge gradually in the subsequent sections.

REFERENCES

LESLIE, P. H. (1945). On the use of matrices in certain population mathematics. *Biometrika*, **33**, 183.

SEAL, H. L. (1945). The mathematics of a population composed of k stationary strata. *Biometrika*, **33**, 226.

A SIMPLE APPROACH TO CONFOUNDING AND FRACTIONAL REPLICATION IN FACTORIAL EXPERIMENTS

By O. KEMPTHORNE, *Rothamsted Experimental Station*

Introduction

The design and analysis of factorial experiments was described in 1937 by Yates in considerable detail. In his treatment Yates described first the 2^n system and then went on to deal with 3^n experiments and experiments of the 2^m3^n type. The 2^n system is capable of very easy explanation, but with experiments of higher order both the design and analysis become of increasing complexity. It is the purpose of this paper to present a general method by which factorial designs of the type p^n may be examined, in respect of both confounding and fractional replication. The method will be described by explanation of the rules for the 2^n and 3^n systems and corresponds quite closely to that given by Fisher (1942). The present approach presents confounding and fractional replication as different aspects of the same process. Experimental designs suggested by Plackett & Burman (1946) are also discussed.

The 2^n system

In this system all combinations of n factors each at two levels are tested. The totality of treatment combinations may be represented by the points of an n-dimensional lattice, each side being of unit length. Let the factors be $x_1, x_2, \ldots, x_n$ and take n mutually orthogonal axes $y_1 \ldots y_n$. The point $(000 \ldots 0)$ will then represent the control treatment, $(1000 \ldots 0)$ the treatment consisting of x_1 at the upper level and all the other factors at the lower level, and so on. The treatment effect of x_1 is the difference of the means of the yields of plots receiving x_1 and those not receiving x_1. It is therefore the difference between the mean of the plots represented by points lying on the plane $y_1 = 1$ and the mean of those represented by the points on the plane $y_1 = 0$. The interaction of x_1 and x_2 is the difference between the means of those plots represented by $y_1 = 1$, $y_2 = 1$ or $y_1 = 0$, $y_2 = 0$ and those represented by $y_1 = 0$, $y_2 = 1$ and $y_1 = 1$, $y_2 = 0$, i.e. the difference of the means of those plots for which

$$y_1 + y_2 = 2 \text{ or } = 0 \pmod 2,$$

and those for which

$$y_1 + y_2 = 1 \pmod 2.$$

Similarly, the triple interaction of x_1, x_2 and x_3 is the difference between the means of those plots for which

$$y_1 + y_2 + y_3 = 0 \pmod 2,$$

and those for which

$$y_1 + y_2 + y_3 = 1 \pmod 2.$$

This process can be continued to the consideration of the interaction of $x_1, x_2, \ldots, x_n$ which is the difference between the mean of those plots for which

$$y_1 + y_2 + y_3 + \ldots + y_n = 0 \pmod 2,$$

and the mean of those for which

$$y_1 + y_2 + y_3 + \ldots + y_n = 1 \pmod 2.$$

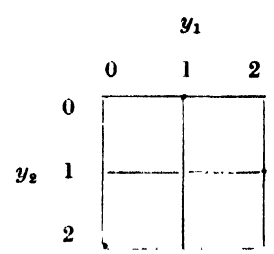
In the n-dimensional space parallel hyper-planes may be drawn containing the points of the lattice, such that the total yield forming the positive part of an interaction is obtained from

a set of parallel hyper-planes equidistant from each other. Likewise the negative part is obtained from another set of parallel hyper-planes, each plane of which lies midway between two planes of the first set.

THE 3^n SYSTEM

With n factors at each of three levels the treatment combinations are given by an n-dimensional lattice, each side being of length two units and containing three points. The treatment contrasts may be described as in the 2^n system with some slight modifications.

Any contrast in the 3^n system involves the comparison of three totals of the yields of 3^{n-1} plots, and may be represented by the comparison of the differences between the yields of the plots lying on three sets of parallel hyper-planes. For example, if $n = 2$ the lattice is as follows:



The main effect of x_1 is the difference between the totals of yields of plots for which $y_1 = 0$, $y_1 = 1$ and $y_1 = 2$. The I component* of the interaction of x_1 and x_2 is the difference between the totals of the yields of plots for which

$$y_1 - y_2 = 0, \quad y_1 - y_2 = 1, \quad \text{and} \quad y_1 - y_2 = 2.$$

The J component is given by the contrast between the yields of plots for which

$$y_1 + y_2 = 0, \quad y_1 + y_2 = 1, \quad \text{and} \quad y_1 + y_2 = 2.$$

Anticipating the extension to cases when n is greater than 2, the equations for the I component may be written as follows:

$$X_1X_2(I_0): \quad y_1 + 2y_2 = 0 \quad (\text{mod } 3),$$
$$X_1X_2(I_1): \quad y_1 + 2y_2 = 1 \quad (\text{mod } 3),$$
$$X_1X_2(I_2): \quad y_1 + 2y_2 = 2 \quad (\text{mod } 3).$$

If x_1 and x_2 (and therefore y_1 and y_2) are interchanged, then $X_2X_1(I_0)$ is given by the equations $y_2 + 2y_1 = 0$, $X_2X_1(I_1)$ by $y_2 + 2y_1 = 1$, $X_2X_1(I_2)$ by $y_2 + 2y_1 = 2$, all mod 3. But the equation $y_2 + 2y_1 = 0$ (mod 3) is identical with the equation $y_1 + 2y_2 = 0$ (mod 3), since $3y_1 + 3y_2 = 0$ (mod 3), whatever the values of y_1 and y_2; $X_2X_1(I_0)$ is therefore equal to $X_1X_2(I_0)$. Subtracting the equation $y_2 + 2y_1 = 1$ (mod 3) from the equation

$$3y_1 + 3y_2 = 0 = 3 \quad (\text{mod } 3),$$

we get $y_1 + 2y_2 = 2$ (mod 3); $X_2X_1(I_1)$ is therefore identical with $X_1X_2(I_2)$. It is obvious from the equations given above for the J component that $X_1X_2(J_i) = X_2X_1(J_i)$ for $i = 0$, 1 and 2.

* Yates's terminology for the components of interactions is used where convenient, but it is more convenient to refer to I_1, I_2 and I_3 as I_0, I_1 and I_2 respectively.

Considering the case $n = 3$, it is easily seen that the second order interaction may be split into four parts each consisting of the contrasts between three totals. These may be represented by the following equations:

$$
\begin{aligned}
&\text{(I)} & y_1 + y_2 + y_3 &= 0 \pmod 3, \\
&& y_1 + y_2 + y_3 &= 1 \\
&& y_1 + y_2 + y_3 &= 2 \\
&\text{(II)} & y_1 + 2y_2 + y_3 &= 0 \pmod 3, \\
&& y_1 + 2y_2 + y_3 &= 1 \\
&& y_1 + 2y_2 + y_3 &= 2 \\
&\text{(III)} & y_1 + y_2 + 2y_3 &= 0 \pmod 3, \\
&& y_1 + y_2 + 2y_3 &= 1 \\
&& y_1 + y_2 + 2y_3 &= 2 \\
&\text{(IV)} & y_1 + 2y_2 + 2y_3 &= 0 \pmod 3. \\
&& y_1 + 2y_2 + 2y_3 &= 1 \\
&& y_1 + 2y_2 + 2y_3 &= 2
\end{aligned}
$$

In order these have been named by Yates

$$Z, \quad X, \quad Y, \quad W.$$

It is interesting in passing to note the relations between Z, X, Y and W for permutations of the order of the factors. It is obvious from (I) that Z is invariant for any change in order of the three factors X_1, X_2 and X_3. Interchanging y_2 and y_3, equations (II) become equations (III), so that $ABC(X) = ACB(Y)$. The following interchanges may be easily verified (using the equation $3y_1 + 3y_2 + 3y_3 = 0 \pmod 3$ where necessary):

$$ABC(X) = BCA(Y) = CAB(Y) = ACB(Y) = CBA(X) = BAC(W).$$

From the equations, it is clear that Z, X and W may be computed in the way given by Yates, since

$$
\begin{aligned}
Z &= J\{x_1, J(x_2, x_3)\}, & Y &= J\{x_1, I(x_2, x_3)\}, \\
X &= I\{x_1, I(x_2, x_3)\}, & W &= I\{x_1, J(x_2, x_3)\},
\end{aligned}
$$

$I(x_2, x_3)$ and $J(x_2, x_3)$ being evaluated for each level of x_1. The extension to the case $n = 4$ is again obvious; the main effects, two-factor and three-factor interactions, follow as in the above, and the four-factor interaction may be split into eight comparisons of three totals:

$$
\begin{aligned}
&\text{I} & y_1 + y_2 + y_3 + y_4 &= 0, 1, 2 \pmod 3, \\
&\text{II} & y_1 + y_2 + y_3 + 2y_4 &= 0, 1, 2 \pmod 3, \\
&\text{III} & y_1 + y_2 + 2y_3 + y_4 &= 0, 1, 2 \pmod 3, \\
&\text{IV} & y_1 + y_2 + 2y_3 + 2y_4 &= 0, 1, 2 \pmod 3, \\
&\text{V} & y_1 + 2y_2 + y_3 + y_4 &= 0, 1, 2 \pmod 3, \\
&\text{VI} & y_1 + 2y_2 + y_3 + 2y_4 &= 0, 1, 2 \pmod 3, \\
&\text{VII} & y_1 + 2y_2 + 2y_3 + y_4 &= 0, 1, 2 \pmod 3, \\
&\text{VIII} & y_1 + 2y_2 + 2y_3 + 2y_4 &= 0, 1, 2 \pmod 3.
\end{aligned}
$$

As in the case of two factors, the effect of permutations of the order on the components of X, Y, Z and W may be easily obtained. The four-factor interactions may be computed by putting the equations given above into the following form:

$$\begin{aligned} \mathrm{I} &= J\{x_1, Z\}, & \mathrm{V} &= I\{x_1, W\}, \\ \mathrm{II} &= J\{x_1, Y\}, & \mathrm{VI} &= I\{x_1, X\}, \\ \mathrm{III} &= J\{x_1, X\}, & \mathrm{VII} &= I\{x_1, Y\}, \\ \mathrm{IV} &= J\{x_1, W\}, & \mathrm{VIII} &= I\{x_1, Z\}, \end{aligned}$$

where the three components of W, X, Y, Z of x_2, x_3, x_4 (in that order) are evaluated for each level of x_1.

The p^n system

The total of $p^n - 1$ degrees of freedom, where p is a prime, in the analysis of variance of a p^n experiment may be split into $(p^n-1)/(p-1)$ sets of $(p-1)$ degrees of freedom, the contrasts being given by the following hyper-planes:

$$\text{Main effects} \quad \begin{array}{l} y_1 = 0, 1, 2, \ldots, p-1, \\ y_2 = 0, 1, 2, \ldots, p-1. \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ y_p = 0, 1, 2, \ldots, p-1. \end{array} \quad (\text{mod } p),$$

$$\begin{array}{r} \text{Interactions of pairs of} \\ \text{factors, e.g. of } x_1 \text{ and } x_2 \end{array} \left\{ \begin{array}{r} y_1 + y_2 = 0, 1, 2, \ldots, p-1, \\ y_1 + 2y_2 = 0, 1, 2, \ldots, p-1, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_1 + (p-1)y_2 = 0, 1, 2, \ldots, p-1. \end{array} \right. \quad (\text{mod } p),$$

and so on to the interaction between all the factors which is given by the hyper-planes

$$a_1 y_1 + a_2 y_2 + a_3 y_3 + \ldots + a_n y_n = 0, 1, 2, \ldots, p-1 \quad (\text{mod } p),$$

where a_1 equals 1 and $a_2, a_3, \ldots, a_n$ each may take all values from 1 to $p-1$.

Simplification of notation

The $p^n - 1$ degrees of freedom in the p^n system may be split into $(p^n-1)/(p-1)$ sets of $(p-1)$ degrees of freedom, given by the above hyper-planes, but it is only necessary to specify one hyper-plane of each set of the parallel hyper-planes.

All the comparisons may be denoted by $y_1^{\alpha_1} y_2^{\alpha_2}, \ldots, y_n^{\alpha_n}$, the symbol meaning that the comparisons are given by the hyper-planes

$$\alpha_1 y_1 + \alpha_2 y_2 + \ldots + \alpha_n y_n = 0, 1, 2, \ldots, p-1 \quad (\text{mod } p).$$

In order to obtain an enumeration which covers all the possibilities once and once only, it is necessary to use the rule that the factors are always written down in ascending order—i.e. $y_i^{\alpha_i} y_j^{\alpha_j} y_k^{\alpha_k}$, etc., such that $i < j < k \ldots$ and that $\alpha_i = 1$.

THE 3^n SYSTEM IN THE REVISED NOTATION

As an example, the 3^3 system will be examined in detail. The effects are represented by y_1, y_2, y_3; interactions between pairs, $y_1y_2, y_1y_2^2, y_1y_3, y_1y_3^2, y_2y_3, y_2y_3^2$; interactions between all three factors, $y_1y_2y_3$, $y_1y_2y_3^2$, $y_1y_2^2y_3$, $y_1y_2^2y_3^2$. Any other combination of powers of the y's can be reduced to the above set.

It is interesting to examine the interactions of the effects and interactions. In the case of the 2^n system, Yates refers to the generalized interaction of two interactions $ABCD$ and CDE say, which is ABE. The interaction of effects or interactions A and B consists of AB and AB^2 in the 3^n system.

(*a*) The interactions of main effects are obviously interactions between pairs of factors.

(*b*) The interactions of main effects and two-factor interactions with one letter in common are two-factor interactions and main effects: e.g. the interactions of y_1 and y_1y_2 are

$$y_1^2y_2 = y_1y_2^2, \quad \text{and} \quad y_1^3y_2^2 = y_2,$$

and the interactions of y_1 and $y_1y_2^2$ are $y_1^2y_2^2 = y_1y_2$, and $y_1^3y_2^4 = y_2$.

(*c*) The interaction between main effects and three-factor interaction are two-factor and three-factor interactions:

Between	Interactions	
y_1 and $y_1y_2y_3$	$y_1^2y_2y_3 = y_1y_2^2y_3^2$,	$y_1^3y_2^2y_3^2 = y_2y_3$
y_1 and $y_1y_2y_3^2$	$y_1^2y_2y_3^2 = y_1y_2^2y_3$,	$y_1^3y_2^2y_3^4 = y_2y_3^2$
y_1 and $y_1y_2^2y_3$	$y_1^2y_2^2y_3 = y_1y_2y_3^2$,	$y_1^3y_2^4y_3^2 = y_2y_3^2$
y_1 and $y_1y_2^2y_3^2$	$y_1^2y_2^2y_3^2 = y_1y_2y_3$,	$y_1^3y_2^4y_3^4 = y_2y_3$

(*d*) The interaction between two-factor interactions are exemplified in the following table:

Between	Interactions	
y_1y_2 and $y_1y_2^2$	$y_1^2y_2^3 = y_1$,	$y_1^3y_2^5 \quad = y_2$
y_1y_2 and y_2y_3	$y_1y_2^2y_3$,	$y_1y_2^3y_3^2 = y_1y_3^2$
y_1y_2 and $y_2y_3^2$	$y_1y_2^2y_3^2$,	$y_1y_2^3y_3^4 = y_1y_3$

(*e*) The interaction between two-factor and three-factor interactions are exemplified in the following table:

Between and	y_1y_2		$y_1y_2^2$	
$y_1y_2y_3$	$y_1y_2y_3^2$	y_3	$y_1y_3^2$	$y_2y_3^2$
$y_1y_2y_3^2$	$y_1y_2y_3$	y_3	y_1y_3	y_2y_3
$y_1y_2^2y_3$	$y_1y_3^2$	y_2y_3	$y_1y_2^2y_3^2$	y_3
$y_1y_2^2y_3^2$	y_1y_3	$y_2y_3^2$	$y_1y_2^2y_3$	y_3

The interactions between two-factor and three-factor interactions are therefore two-factor interactions in some cases and main effects and three-factor interactions in the other cases.

(*f*) The interactions between three-factor interactions are set out in the following diagram:

Between \ and	$y_1y_2y_3$	$y_1y_2y_3^2$	$y_1y_2^2y_3$	$y_1y_2^2y_3^2$
$y_1y_2y_3$	— —	y_1y_2, y_3	y_1y_3, y_2	y_1, y_2y_3
$y_1y_2y_3^2$		— —	y_1, $y_2y_3^2$	$y_1y_3^2$, y_2
$y_1y_2^2y_3$			— —	$y_1y_2^2$, y_3
$y_1y_2^2y_3^2$				— —

Confounding

Confounding or the allocation of treatment combinations to blocks implies the allocation of all the points of the lattice into p^c sets, of p^{n-c} points, such that the comparisons between these sets involve particular sets of $p-1$ degrees of freedom. The aim of confounding is to reduce the effect of soil heterogeneity by reducing block size, but ensuring that the block comparisons have little possible practical importance.

If comparisons $A = y_1^{\alpha_1}y_2^{\alpha_2}\dots y_n^{\alpha_n}$ and $B = y_1^{\beta_1}y_2^{\beta_2}\dots y_n^{\beta_n}$ are confounded, then so is their generalized interaction, i.e. all the products of these two, i.e. AB, $AB^2, \dots, AB^{p-1}$. For, if the treatment combinations for which $\alpha_1y_1+\alpha_2y_2+\dots+\alpha_ny_n$ is equal to $0, 1, 2, \dots, p-1$ are put into separate blocks and also those treatments for which $\beta_1y_1+\beta_2y_2+\dots+\beta_ny_n$ is equal to $0, 1, 2, \dots, p-1$, then $(\alpha_1+\lambda\beta_1)y_1+(\alpha_2+\lambda\beta_2)y_2+\dots+(\alpha_n+\lambda\beta_n)y_n$ is equal (mod p) to $0, 1, 2, \dots, p-1$ for all λ from 0 to $p-1$.

The present approach to confounding of the 2^n system is identical with that given by Yates and we proceed to consider the rather more complex case of the 3^n system.

(*a*) 3^3 *system*

(1) *In blocks of* 3^2. Any three-factor interaction may be confounded.

(2) *In blocks of* 3. We cannot confine the confounded degrees of freedom to three-factor interactions because the generalized interaction of any two reduces to a two-factor interaction and a main effect. If two three-factor interactions, $y_1y_2y_3$ and $y_1y_2^2y_3$ are confounded, the 8 degrees of freedom for blocks may be described as follows:

	D.F.
y_2	2
y_1y_3	2
$y_1y_2y_3$	2
$y_1y_2^2y_3$	2
	8

We can, however, choose three two-factor interactions and one three-factor interaction pair for our block comparisons.

(b) 3^4 system in blocks of 3^2

It is immediately obvious that we can confound two two-factor interactions and two higher-order interaction pairs to give blocks of nine. The important point, however, is to find a design confounding only three-factor interaction pairs.

We therefore evaluate the interactions of all pairs of three-factor interactions, which have two letters in common. These may be derived from the interaction of $y_1y_2y_3$ with the four three-factor interactions of y_1, y_2 and y_4, which are as follows:

Interaction of $y_1y_2y_3$	and	$y_1y_2y_4$	$y_1y_2y_3^2y_4^2$	and	$y_3y_4^2$,
Interaction of $y_1y_2y_3$	and	$y_1y_2y_4^2$	$y_1y_2y_3^2y_4$	and	y_3y_4,
Interaction of $y_1y_2y_3$	and	$y_1y_2^2y_4$	$y_1y_3^2y_4^2$	and	$y_2y_3^2y_4$,
Interaction of $y_1y_2y_3$	and	$y_1y_2^2y_4^2$	$y_1y_3^2y_4$	and	$y_2y_3^2y_4^2$.

Obviously there are many designs for the 3^4 design in nine blocks of nine plots confounding three-factor interactions. Those which confound four-factor interactions must also confound two-factor interactions. The names of the confounded interactions and their squares (each of which corresponds to the same grouping as the element itself) form a group with the identity and the equation $y_i^3 = 1$, for all i, and further work is presumably most promising on these lines.

(c) 3^5 in blocks of 9

There is no design confounding only three-factor or higher-order interactions. If one two-factor interaction can be sacrificed, a possible scheme of confounding is given by the following table of generalized interactions:

Between \ and	$y_1y_2y_3$	$y_1y_2^2y_4^2$	$y_1y_3^2y_4$	$y_2y_3^2y_4^2$
$y_3y_4y_5$	$y_1y_2y_3^2y_4y_5$ $y_1y_2y_4^2y_5^2$	$y_1y_2^2y_3y_5$ $y_1y_2^2y_3^2y_4y_5^2$	$y_1y_4^2y_5$ $y_1y_3y_5^2$	y_2y_5 $y_2y_3y_4y_5^2$

This two-factor interaction is estimated by the comparison of three sets of nine blocks, and the accuracy of the estimate will be low.

(d) 3^6 in blocks of 27

We may, for example, confound the following:

	$y_2y_4y_6$	
$y_1y_2y_3$	$y_1y_2^2y_3y_4y_6$	$y_1y_3y_4^2y_6^2$
$y_1y_4y_5$	$y_1y_2y_4^2y_5y_6$	$y_1y_2^2y_5y_6^2$
$y_1y_2^2y_3^2y_4^2y_5^2$	$y_1y_3^2y_5^2y_6$	$y_1y_2y_3^2y_4y_5^2y_6^2$
$y_2y_3y_4^2y_5^2$	$y_2y_3^2y_5y_6^2$	$y_3y_4y_5^2y_6^2$

Three three-factor interactions, six four-factor interactions, three five-factor interactions and one six-factor interaction are confounded. If y_6 is omitted from all the above expressions

we obtain a 3^5 experiment in blocks of nine confounding one two-factor interaction, seven three-factor interactions, three four-factor and two five-factor interactions—that is, the design given above for the 3^5 system.

Extension to more complicated cases

Extensions of the above to more complicated cases should most easily be achieved by the use of group theory. The confounding of a p^n design in p^c blocks corresponds to a group of $\frac{1}{2}(p^c+1)$ elements such that all except the unit element involve at least a certain number of letters. For most agricultural experiments each element should contain at least three letters, so that no main effects or two-factor interactions are confounded. The group is an Abelian group and if A and B are elements of the group so are $AB, AB^2, \ldots, AB^{p-1}$. The order of each element is p, and if A is an element so are the first $(p-1)$ powers of A. This aspect is being followed, and it is hoped will yield results.

Fractional replication in the 2^n system

Some principles of fractional replication have been worked out over the past few years at Rothamsted (Finney, 1945). In the case of a 2^n system, with factors $a_1 \ldots a_n$ say, a half-replicate might consist of those treatment combinations which form the positive part of the interaction $A_1 A_2 \ldots A_n$. Each function of the plot yields consisting of the sum of one-half of them minus the sum of the other half then corresponds to two degrees of freedom. Alternatively, each degree of freedom has one alias, and the aim in fractional replication is to design the experiment so that the aliases of effects which the experimenter wishes to measure are high-order interactions which could not possibly have practical significance.

For convenience of presentation, we develop first the theory for the case of the 2^n system. Suppose that of all the points on the lattice for the 2^n system, only those points for which

$$y_1 + y_2 + y_3 + \ldots + y_n = 0$$

are included in the experiment. Then the points on the hyper-plane $y_1 = 0$, also lie on the plane $y_2 + y_3 + \ldots + y_n = 0$, and likewise those for which $y_1 = 1$ lie on the plane

$$y_2 + y_3 + \ldots + y_n = 1.$$

The contrast which we have denoted by y_1 is therefore identical with that denoted by $y_2 y_3 y_4 \ldots y_n$. Again, if we suppose that only those treatment combinations are tested which lie on the hyper-planes

$$\alpha_1 y_1 + \alpha_2 y_2 + \ldots + \alpha_n y_n = 0, \quad \beta_1 y_1 + \beta_2 y_2 + \ldots + \beta_n y_n = 0,$$

then the points will also lie on the intersection of these planes which is given by the equation

$$(\alpha_1 + \beta_1) y_1 + (\alpha_2 + \beta_2) y_2 + \ldots + (\alpha_n + \beta_n) = 0 \pmod 2.$$

The points which lie on the hyper-planes

$$\gamma_1 y_1 + \gamma_2 y_2 + \ldots + \gamma_n y_n = 0, 1 \pmod 2$$

will also lie on the planes

$$\begin{aligned}
(\alpha_1 + \gamma_1) y_1 + (\alpha_2 + \gamma_2) y_2 \quad &+ \ldots + \quad (\alpha_n + \gamma_n) y_n = 0, 1 \pmod 2, \\
(\beta_1 + \gamma_1) y_1 + (\beta_2 + \gamma_2) y_2 \quad &+ \ldots + \quad (\beta_n + \gamma_n) y_n = 0, 1 \pmod 2, \\
(\alpha_1 + \beta_1 + \gamma_1) y_1 + (\alpha_2 + \beta_2 + \gamma_2) y_2 &+ \ldots + (\alpha_n + \beta_n + \gamma_n) y_n = 0, 1 \pmod 2.
\end{aligned}$$

Changing to the simpler notation, these results may be obtained by equating to unity the symbols corresponding to the effects which the experiment cannot measure (as only treat-

ment combinations of the same sign in the function giving the effect are included) and multiplying the symbol corresponding to a particular effect by these symbols. Thus we put

$$I = y_1^{\alpha_1} y_2^{\alpha_2} \dots y_n^{\alpha_n} = y_1^{\beta_1} y_2^{\beta_2} \dots y_n^{\beta_n} = y_1^{\alpha_1+\beta_1} y_2^{\alpha_2+\beta_2} \dots y_n^{\alpha_n+\beta_n},$$

then the contrast

$$y_1^{\gamma_1} y_2^{\gamma_2} \dots y_n^{\gamma_n}$$

is the same as those given by

$$y_1^{\alpha_1+\gamma_1} y_2^{\alpha_2+\gamma_2} \dots y_n^{\alpha_n+\gamma_n}, \quad y_1^{\beta_1+\gamma_1} y_2^{\beta_2+\gamma_2} \dots y_n^{\beta_n+\gamma_n} \quad \text{and} \quad y_1^{\alpha_1+\beta_1+\gamma_1} y_2^{\alpha_2+\beta_2+\gamma_2} \dots y_n^{\alpha_n+\beta_n+\gamma_n},$$

where each power is reduced modulus 2.

2^n SYSTEM WITHOUT SUBDIVISION INTO BLOCKS

We now consider some of the possibilities of partial replication for the 2^n system. The basis of designs with fractional replication is the choice of an identity relationship; most of the possible relationships are of no value, and we consider only those which yield the least possible confusion between main effects and first-order interactions.

Half-replication

$n = 3$. If we take $I = y_1 y_2 y_3$, then $y_1 = y_1(y_1 y_2 y_3) = y_1^2 y_2 y_3 = y_2 y_3$. Such a design which confuses main effects and two-factor interactions would not be of any practical use.

$n = 4$. If we take $I = y_1 y_2 y_3 y_4$, then the aliases are exemplified by

$$y_1 = y_2 y_3 y_4 \quad \text{and} \quad y_1 y_2 = y_3 y_4.$$

Such a design would not be used unless the experimenter were confident that two-factor interactions were negligible.

$n = 5$. If we take $I = y_1 y_2 y_3 y_4 y_5$, then the aliases are exemplified by

$$y_1 = y_2 y_3 y_4 y_5 \quad \text{and} \quad y_1 y_2 = y_3 y_4 y_5.$$

A half-replicate with five or more factors is feasible when there is no necessity to remove heterogeneity by the use of blocks, since main effects will have aliases which are interactions of four factors at least, and two-factor interactions will have aliases which are interactions of at least three factors.

Quarter-replication

Each degree of freedom will now have three aliases. For each value of n we give the identity relationship and typical alias relationships.

$n = 4$. $\quad I = y_1 y_2 = y_3 y_4 = y_1 y_2 y_3 y_4;$

then $\quad y_1 = y_2 = y_1 y_3 y_4 = y_2 y_3 y_4$ and $y_1 y_3 = y_2 y_3 = y_1 y_4 = y_2 y_4.$

$n = 5$. $\quad I = y_1 y_2 = y_3 y_4 y_5 = y_1 y_2 y_3 y_4 y_5 \quad (a),$

or $\quad I = y_1 y_2 y_3 = y_3 y_4 y_5 = y_1 y_2 y_4 y_5 \quad (b).$

(*a*) Gives $\quad y_1 = y_2 = y_1 y_3 y_4 y_5 = y_2 y_3 y_4 y_5$ and $y_1 y_3 = y_2 y_3 = y_1 y_4 y_5 = y_2 y_4 y_5.$

(*b*) Gives $\quad y_1 = y_2 y_3 = y_1 y_3 y_4 y_5 = y_2 y_4 y_5.$

$n = 6$. $\quad I = y_1 y_2 y_3 y_4 = y_3 y_4 y_5 y_6 = y_1 y_2 y_5 y_6;$

then $\quad y_1 = y_2 y_3 y_4 = y_1 y_3 y_4 y_5 y_6 = y_2 y_5 y_6$

and $\quad y_1 y_2 = y_3 y_4 = y_1 y_2 y_3 y_4 y_5 y_6 = y_5 y_6.$

$n = 7$. $\quad I = y_1 y_2 y_3 y_4 = y_4 y_5 y_6 y_7 = y_1 y_2 y_3 y_5 y_6 y_7;$

then $\quad y_1 = y_2 y_3 y_4$ and $y_1 y_2 = y_3 y_4.$

$n = 8$. $\quad I = y_1 y_2 y_3 y_4 y_5 = y_4 y_5 y_6 y_7 y_8 = y_1 y_2 y_3 y_6 y_7 y_8;$

then $\quad y_1 = y_2 y_3 y_4 y_5$ and $y_1 y_2 = y_3 y_4 y_5.$

Designs in quarter replicate are therefore possible when n is greater than or equal to 8.

High-order fractional replication

In general, the existence of fractional designs of the 2^n system with fraction 2^p, which will be useful where information on all main effects and two-factor interactions is required, depends on the existence of a group of 2^p elements, one element being unity and the other elements all containing at least five letters. No simple method has been found of enumerating such groups, but it is perhaps worth recording the following designs which appear to represent the greatest degree of fractional replication possible.

(*a*) *Eighth replication*

If we are testing ten or more factors at each of two levels, one-eighth of a replication will enable main effects and two-factor interactions to be estimated. An appropriate identity relationship is the following:

$$\begin{aligned} I &= y_1y_2y_3y_4y_5 = y_1y_2y_6y_7y_8 = y_3y_4y_5y_6y_7y_8 \\ &= y_1y_3y_7y_9y_{10} = y_2y_4y_5y_7y_9y_{10} = y_2y_3y_6y_8y_9y_{10} = y_1y_4y_5y_6y_8y_9y_{10}. \end{aligned}$$

Thus ten main effects and forty-five two-factor interactions may be estimated from a trial testing 128 of the 1024 possible treatment combinations.

(*b*) *Sixteenth replication*

If we are testing twelve or more factors a possible identity relationship is the following:

$$\begin{aligned} I &= y_1y_2y_3y_4y_5 = y_1y_2y_6y_7y_8 = y_3y_4y_5y_6y_7y_8 = y_1y_2y_9y_{10}y_{11} \\ &= y_3y_4y_5y_9y_{10}y_{11} = y_6y_7y_8y_9y_{10}y_{11} = y_1y_2y_3y_4y_5y_6y_7y_8y_9y_{10}y_{11} \\ &= y_1y_3y_6y_9y_{12} = y_2y_4y_5y_6y_9y_{12} = y_2y_3y_7y_8y_9y_{12} = y_1y_4y_5y_7y_8y_9y_{12} \\ &= y_2y_3y_6y_{10}y_{11}y_{12} = y_1y_4y_5y_6y_{10}y_{11}y_{12} = y_1y_3y_7y_8y_{10}y_{11}y_{12} = y_2y_4y_5y_7y_8y_{10}y_{11}y_{12}. \end{aligned}$$

In this case twelve main effects and sixty-six two-factor interactions may be estimated from a trial testing 256 of the possible 4096 treatment combinations.

The extent to which these designs will be of practical value depends very much on the existence of a sufficient mass of reasonably homogeneous material to test the large number of treatment combinations without the necessity of dividing the material into smaller batches and using the device of confounding. An experiment involving say 256 different treatment combinations is not large by modern standards. At Rothamsted, for example, an experiment involving 200 distinct treatments on 300 plots has been carried out for some years: this experiment was, however, made possible by utilizing the elimination of the effects of soil heterogeneity by highly complex confounding; the design, in fact, consisted of three 5×5 lattice squares necessitating seventy-five plots, and each of these plots was split into four subplots. The advantages of testing twelve factors, say, at the same time under virtually the same experimental conditions cannot, however, be ignored. Such an experiment should have more value, other things being equal, than two distinct experiments each testing some of the factors. An examination has not been made of the possibilities of reducing block size by confounding for the above two designs, but it is probably necessary to sacrifice a few two-factor interactions.

The relationship between fractional replication and confounding

It is clear that fractional replication and confounding are different aspects of the same process. A 2^n design of 2^p blocks may be described as a 1 in 2^p replicate of a 2^{n+p} design with no subdivision into blocks, by regarding the blocks as a 2^p system in p factors. As an

example, consider the 2^5 design in y_1, y_2, y_3, y_4 and y_5 laid out in four blocks of eight and confounding $y_1y_2y_3$, $y_3y_4y_5$ and $y_1y_2y_4y_5$; superimposing two pseudo-factors b_1 and b_2, the experiment is a quarter-replicate of a 2^7 design in $y_1, y_2, y_3, y_4, y_5, b_1, b_2$. The identity on which the quarter replicate is based is given by the equations

$$b_1 = y_1y_2y_3, \quad b_2 = y_3y_4y_5, \quad b_1b_2 = y_1y_2y_4y_5$$

or the equation

$$I = y_1y_2y_3b_1 = y_3y_4y_5b_2 = y_1y_2y_4y_5b_1b_2.$$

If we examine this equation in the same way as in the previous sections, we find that the design depends on the fact that the aliases of the following type may be ignored:

$$y_1 = y_2y_3b_1 = y_1y_3y_4y_5b_2 = y_2y_4y_5b_1b_2,$$

$$y_1y_2 = y_3b_1 = y_1y_2y_3y_4y_5b_2 = y_4y_5b_1b_2.$$

This example is worth pursuing. The design is frequently used with one replication only, the error being estimated from three-factor and higher-order interactions. We set out below the identity and 31 degrees of freedom together with all their aliases and their usual place in the analysis of variance—blocks (B), treatment (T), or error (E). For convenience of printing we denote the factors tested in the experiment by a, b, c, d, e instead of $y_1y_2y_3y_4y_5$ and the block factors by x and y. Capitals are used for treatment effects thus conforming to present usage.

I	$= ABCX$	$= CDEY$	$= ABDEXY$	
A	$= BCX$	$= ACDEY$	$= BDEXY$	T
B	$= ACX$	$= BCDEY$	$= ADEXY$	T
AB	$= CX$	$= ABCDEY$	$= DEXY$	T
C	$= ABX$	$= DEY$	$= ABCDEXY$	T
AC	$= BX$	$= ADEY$	$= BCDEXY$	T
BC	$= AX$	$= BDEY$	$= ACDEXY$	T
ABC	$= X$	$= ABDEY$	$= CDEXY$	B
D	$= ABCDX$	$= CEY$	$= ABEXY$	T
AD	$= BCDX$	$= ACEY$	$= BEXY$	T
BD	$= ACDX$	$= BCEY$	$= AEXY$	T
ABD	$= CDX$	$= ABCEY$	$= EXY$	E
CD	$= ABDX$	$= EY$	$= ABCEXY$	T
ACD	$= BDX$	$= AEY$	$= BCEXY$	E
BCD	$= ADX$	$= BEY$	$= ACEXY$	E
$ABCD$	$= DX$	$= ABEY$	$= CEXY$	E
E	$= ABCEX$	$= CDY$	$= ABDXY$	T
AE	$= BCEX$	$= ACDY$	$= BDXY$	T
BE	$= ACEX$	$= BCDY$	$= ADXY$	T
ABE	$= CEX$	$= ABCDY$	$= DXY$	E
CE	$= ABEX$	$= DY$	$= ABCDXY$	T
ACE	$= BEX$	$= ADY$	$= BCDXY$	E
BCE	$= AEX$	$= BDY$	$= ACDXY$	E
$ABCE$	$= EX$	$= ABDY$	$= CDXY$	E
DE	$= ABCDEX$	$= CY$	$= ABXY$	T
ADE	$= BCDEX$	$= ACY$	$= BXY$	E
BDE	$= ACDEX$	$= BCY$	$= AXY$	E
$ABDE$	$= CDEX$	$= ABCY$	$= XY$	B
CDE	$= ABDEX$	$= Y$	$= ABCXY$	B
$ACDE$	$= BDEX$	$= AY$	$= BCXY$	E
$BCDE$	$= ADEX$	$= BY$	$= ACXY$	E
$ABCDE$	$= DEX$	$= ABY$	$= CXY$	E

If we take for each linear function of the yields the alias involving the smallest possible number of letters, but remembering that x, y are pseudo-factors, so that X, Y and XY are of

equal importance and therefore XY should be regarded as a main effect and not an interaction, we have the following allocation of contrasts to the three components of the analysis of variance:

Blocks:	$X,\ Y,\ XY.$
Treatments:	$A,\ B,\ C,\ D,\ E.$
	$AB = CX,\quad AC = BX,\quad BC = AX,$
	$CD = EY,\quad DE = CY,\quad CE = DY.$
	$AD,\ BD,\ AE,\ BE.$
Error:	$AY,\ BY,\ DX,\ EX,\ AXY,\ BXY,\ CXY,\ DXY,\ EXY,\ ACD,\ BCD,\ ACE,\ BCE.$

The four three-factor interactions could equally well be regarded as interactions between two-factor interactions and blocks. It would be anticipated that these would be smaller than the interactions of main effects and blocks. The purpose of the present exposition is to give a clear statement of the possible interpretations of the results of an individual experiment. Further remarks on the problem of interpretation are postponed to a later section in the paper.

An example of fractional replication with confounding

A design which has proved of practical utility is the half-replicate of a 2^6 experiment arranged in four blocks of eight plots.

Call the factors $y_1, y_2, y_3, y_4, y_5, y_6$. Then the best confounding is that in which, using full replication, the block differences are all third-order interactions, say

$$y_1y_2y_3y_4,\quad y_3y_4y_5y_6 \quad\text{and}\quad y_1y_2y_5y_6.$$

But it is impossible to keep main effects and interactions clear with this confounding, whatever interaction is equated to the identity.

If we take the confounded interactions to be of the type

$$y_1y_2y_3,\quad y_3y_4y_5,\quad y_1y_2y_4y_5,$$

and the interaction $y_1y_2y_3y_4y_5y_6$ to be unity, then the following interactions are also confounded:

$$y_4y_5y_6,\quad y_1y_2y_6 \quad\text{and}\quad y_3y_6.$$

It will be found by enumeration of the possibilities that one first-order interaction must be sacrificed. All main effects and the other first-order interactions will have high-order aliases.

It is interesting to examine this design in the same way as the 2^5 above for the relations between block-treatment interactions and treatment interactions.

There are, in fact, only thirty-two independent contrasts, and it is simplest to enumerate these by operating on the identity relationship with the thirty-two possibilities for the 2^5 system omitting y_6. As before, we insert block pseudo-factors. For simplicity of printing we use A, B, C, D, E, F for the factors and X, Y for the block factors. Then

$$I = ABCDEF,\quad X = ABC,\quad Y = CDE,\quad XY = ABDE,$$

and combining these into one relationship, we have

$$I = ABCDEF = ABCX = CDEY = ABDEXY = DEFX = ABFY = CFXY.$$

A complete table of the aliases for this design follows:

I	$=ABCDEF$	$=ABCX$	$=DEFX$	$=CDEY$	$=ABFY$	$=ABDEXY$	$=CFXY$	
A	$=BCDEF$	$=BCX$	$=ADEFX$	$=ACDEY$	$=BFY$	$=BDEXY$	$=ACFXY$	T
B	$=ACDEF$	$=ACX$	$=BDEFX$	$=BCDEY$	$=AFY$	$=ADEXY$	$=BCFXY$	T
AB	$=CDEF$	$=CX$	$=ABDEFX$	$=ABCDEY$	$=FY$	$=DEXY$	$=ABCFXY$	T
C	$=ABDEF$	$=ABX$	$=CDEFX$	$=DEY$	$=ABCFY$	$=ABCDEXY$	$=FXY$	T
AC	$=BDEF$	$=BX$	$=ACDEFX$	$=ADEY$	$=BCFY$	$=BCDEXY$	$=AFXY$	T
BC	$=ADEF$	$=AX$	$=BCDEFX$	$=BDEY$	$=ACFY$	$=ACDEXY$	$=BFXY$	T
ABC	$=DEF$	$=X$	$=ABCDEFX$	$=ABDEY$	$=CFY$	$=CDEXY$	$=ABFXY$	B
D	$=ABCEF$	$=ABCDX$	$=EFX$	$=CEY$	$=ABDFY$	$=ABEXY$	$=CDFXY$	T
AD	$=BCEF$	$=BCDX$	$=AEFX$	$=ACEY$	$=BDFY$	$=BEXY$	$=ACDFXY$	T
BD	$=ACEF$	$=ACDX$	$=BEFX$	$=BCEY$	$=ADFY$	$=AEXY$	$=BCDFXY$	T
ABD	$=CEF$	$=CDX$	$=ABEFX$	$=ABCEY$	$=DFY$	$=EXY$	$=ABCDFXY$	E
CD	$=ABEF$	$=ABDX$	$=CEFX$	$=EY$	$=ABCDFY$	$=ABCEXY$	$=DFXY$	T
ACD	$=BEF$	$=BDX$	$=ACEFX$	$=AEY$	$=BCDFY$	$=BCEXY$	$=ADFXY$	E
BCD	$=AEF$	$=ADX$	$=BCEFX$	$=BEY$	$=ACDFY$	$=ACEXY$	$=BDFXY$	E
$ABCD$	$=EF$	$=DX$	$=ABCEFX$	$=ABEY$	$=CDFY$	$=CEXY$	$=ABDFXY$	T
E	$=ABCDF$	$=ABCEX$	$=DFX$	$=CDY$	$=ABEFY$	$=ABDXY$	$=CEFXY$	T
AE	$=BCDF$	$=BCEX$	$=ADFX$	$=ACDY$	$=BEFY$	$=BDXY$	$=ACEFXY$	T
BE	$=ACDF$	$=ACEX$	$=BDFX$	$=BCDY$	$=AEFY$	$=ADXY$	$=BCEFXY$	T
ABE	$=CDF$	$=CEX$	$=ABDFX$	$=ABCDY$	$=EFY$	$=DXY$	$=ABCEFXY$	E
CE	$=ABDF$	$=ABEX$	$=CDFX$	$=DY$	$=ABCEFY$	$=ABCDXY$	$=EFXY$	T
ACE	$=BDF$	$=BEX$	$=ACDFX$	$=ADY$	$=BCEFY$	$=BCDXY$	$=AEFXY$	E
BCE	$=ADF$	$=AEX$	$=BCDFX$	$=BDY$	$=ACEFY$	$=ACDXY$	$=BEFXY$	E
$ABCE$	$=DF$	$=EX$	$=ABCDFX$	$=ABDY$	$=CEFY$	$=CDXY$	$=ABEFXY$	T
DE	$=ABCF$	$=ABCDEX$	$=FX$	$=CY$	$=ABDEFY$	$=ABXY$	$=CDEFXY$	T
ADE	$=BCF$	$=BCDEX$	$=AFX$	$=ACY$	$=BDEFY$	$=BXY$	$=ACDEFXY$	E
BDE	$=ACF$	$=ACDEX$	$=BFX$	$=BCY$	$=ADEFY$	$=AXY$	$=BCDEFXY$	E
$ABDE$	$=CF$	$=CDEX$	$=ABFX$	$=ABCY$	$=DEFY$	$=XY$	$=ABCDEFXY$	B
CDE	$=ABF$	$=ABDEX$	$=CFX$	$=Y$	$=ABCDEFY$	$=ABCXY$	$=DEFXY$	B
$ACDE$	$=BF$	$=BDEX$	$=ACFX$	$=AY$	$=BCDEFY$	$=BCXY$	$=ADEFXY$	T
$BCDE$	$=AF$	$=ADEX$	$=BCFX$	$=BY$	$=ACDEFY$	$=ACXY$	$=BDEFXY$	T
$ABCDE$	$=F$	$=DEX$	$=ABCFX$	$=ABY$	$=CDEFY$	$=CXY$	$=ABDEFXY$	T

The partition of the degrees of freedom in the analysis of variance which would generally be made is the following:

		D.F.
Blocks		3
Treatments:	Main effects	6
	Interactions	14
Error		8
		31

The table of aliases is condensed below by the omission of all aliases involving more than two factors—counting, as before, XY as a single factor as well as X and Y.

Effects A, B, D, E have aliases of at least three letters, but $C = FXY$ and $F = CXY$.

Effects AD, BD, AE, BE have aliases of at least three letters, but

$$AB = CX = FY, \quad AC = BX, \quad BC = AX, \quad CD = EY, \quad EF = DX,$$

$$CE = DY, \quad DE = FX = CY, \quad DF = EX, \quad BF = AY, \quad AF = BY.$$

In an experiment in which block-treatment interactions cannot be assumed to be negligible in relation to the effects it is desired to estimate, the interpretation of most two-factor interactions is difficult if not impossible. The following identities of practical interest exist for the terms which would be used to estimate the error: ACD, BCD, ACE, BCE have aliases of three letters and are either three-factor interactions or interactions between blocks and two-factor interactions, but $ABD = EXY$, $ABE = DXY$, $ADE = BXY$, and $BDE = AXY$.

This design is very similar in result to the fully replicated but confounded 2^5 design described above.

Fractional replication in the 3^n system

Here we have to consider treatment effects assessed from powers of one-third of a complete replicate. Only those treatment combinations represented by points of the lattice lying on the hyperplane

$$\alpha_1 y_1 + \alpha_2 y_2 + \ldots + \alpha_n y_n = 0, \text{ or } 1, \text{ or } 2 \pmod 3$$

will be included in a one-third replicate.

A particular treatment effect is given by the differences between the means of those plots represented by points on the following three planes:

$$\beta_1 y_1 + \beta_2 y_2 + \ldots + \beta_n y_n = 0 \pmod 3,$$
$$\beta_1 y_1 + \beta_2 y_2 + \ldots + \beta_n y_n = 1 \pmod 3,$$
$$\beta_1 y_1 + \beta_2 y_2 + \ldots + \beta_n y_n = 2 \pmod 3.$$

It is obvious that the points lying on the first plane will also lie on the planes

$$(\beta_1 + \lambda\alpha_1) y_1 + (\beta_2 + \lambda\alpha_2) y_2 + \ldots + (\beta_n + \lambda\alpha_n) y_n = 0 \pmod 3, \quad \text{for } \lambda = 1 \text{ and } 2;$$

the points on the other two planes will lie on these planes with 1 and 2 respectively on the right-hand side of the equation.

The aliases of each pair of degrees of freedom are therefore obtained by multiplication of its symbol by

$$y_1^{\alpha_1} y_2^{\alpha_2} \ldots y_n^{\alpha_n},$$

and by its square.

As an example, suppose a third replicate of a 3^3 design is based on the inclusion only of those treatment combinations represented by the symbol $y_1 y_2 y_3 (y_1 + y_2 + y_3 = 0$ say), then the aliases are exemplified by the relationship $y_1 = y_1 y_2^2 y_3^2 = y_2 y_3$.

The confounding of one replicate of a 3^3 experiment in three blocks of nine plots

A frequently used design is the 3^3 in three blocks of nine plots, testing all combinations of three factors each at three levels. This design is formally a one-third replicate of a 3^4 design. Suppose the factors are y_1, y_2, and y_3 and let blocks be denoted by the pseudo-factor b; a three-factor interaction of y_1, y_2, and y_3, say $y_1 y_2 y_3$, is usually confounded in order to keep main effects and first-order interactions free of block effects.

Then $b = y_1 y_2 y_3$ or $I = y_1 y_2 y_3 b^2$, since $b^3 = 1$.

As in the case of the 2^5 design, we work out the aliases of each pair of degrees of freedom: each pair of degrees of freedom will in this case have two aliases:

$$\begin{array}{llll}
y_1 = y_1 y_2^2 y_3^2 b = y_2 y_3 b^2 & \qquad & y_2 y_3 = y_1 y_2^2 y_3^2 b^2 = y_1 b^2 \\
y_2 = y_1 y_2^2 y_3 b_2^2 = y_1 y_3 b^2 & & y_2 y_3^2 = y_1 y_2^2 b^2 = y_1 y_3^2 b^2 \\
y_3 = y_1 y_2 y_3^2 b^2 = y_1 y_2 b^2 & & y_1 y_2 y_3 = y_1 y_2 y_3 b = b^2 \\
y_1 y_2 = y_1 y_2 y_3^2 b = y_3 b^2 & & y_1 y_2 y_3^2 = y_1 y_2 b = y_3 b \\
y_1 y_2^2 = y_1 y_3^2 b = y_2 y_3^2 b & & y_1 y_2^2 y_3 = y_1 y_3 b = y_2 b \\
y_1 y_3 = y_1 y_2^2 y_3 b = y_2 b^2 & & y_1 y_2^2 y_3^2 = y_1 b = y_2 y_3 b \\
y_1 y_3^2 = y_1 y_2^2 b = y_2 y_3^2 b^2 & & &
\end{array}$$

Here again the identities could result in difficulty in interpretation—as of course could have been predicted from the examination of the possible arrangements in blocks of nine of the 3^4 design. The main effects may be regarded as clear, and three of the first-order interactions. The remaining two-factor interactions could be ascribed to differential effects of the factors on the three blocks. The three-factor interactions which are not confounded with blocks are also ascribable to interactions of main effects and blocks and may therefore be used to form an estimate of the error of these effects.

GENERAL REMARKS ON CONFOUNDING

The device of confounding is used almost without exception in agricultural experiments in order to reduce the block size to twelve or less plots. As the above results indicate there are two aspects which then need careful consideration, (*a*) the estimation of interactions, and (*b*) the estimation of the experimental error.

The main purpose of the factorial design is the estimation of main effects and interactions between pairs of factors and thence of the effect of any one factor in the presence and absence of each of the other factors. It is clear that when it is necessary to remove soil heterogeneity by confounding, the interpretation of a small experiment involving a few factors may be exceedingly difficult because of the possibility of block-treatment interactions. It is possible to use the rule that a large contrast should be regarded as the interaction between whichever pair of main effects is the larger, but this rule will break down in some cases when, for example, the contrast has two aliases AB and CD, and effects A and C are large and B and D small. In the case of a series of experiments, a device which might be helpful is the use of permutations of the possible identity relationships, one at each centre. The modern emphasis in agricultural experimentation is on series of experiments at various places and in several years, rather than on individual experiments. Interactions of pairs of factors will be estimated correctly from a large series of experiments if treatments are assigned at random to blocks.

The evaluation of two-factor interactions for individual experiments depends on the assumption that block-treatment interactions are small compared with the experimental error. Yates (1935) examined several experiments for the existence of such interactions and found no evidence of them. Since that time a large number of experimental results which can be used to provide information on the question have been accumulated, and an investigation of these has indicated that block-treatment interactions are negligible and may be ignored (Kempthorne, 1947).

With regard to the estimation of error, in so far as tests of significance are of interest, it can be said that the analysis of variance does provide a test of significance of the hypothesis that the treatments have an overall effect different from zero. In agricultural experimentation, the term error is used to denote block-treatment interactions. Thus in the simple randomized block experiment, it is possible to evaluate the difference between two treatments from each block, and it is the variability of this difference from block to block which is regarded as the error. In general, as there are usually few blocks, and the error of each comparison would be determined with poor accuracy, the errors of all the possible independent comparisons are pooled to give a common estimate. If the treatments were duplicated at random

within each block, the analysis would be of the form (r being the number of blocks and t of treatments):

	D.F.
Blocks	$r-1$
Treatments	$t-1$
Treatments by blocks	$(r-1)(t-1)$
Within blocks	rt
	$2rt-1$

The component 'within blocks' could more accurately be described as experimental error, but would not be used to evaluate the errors of treatment effects, since the experimenter is interested in the constancy of treatment effects from block to block. There is therefore little point in actually carrying out such an experiment. In a factorial experiment with replication, the components which could be evaluated consist of replicates, effects and low-order interactions, high-order interactions, and interactions of treatments and replicates. On the assumption that the sum of squares for interaction of treatments and replicates is homogeneous, the mean square for high-order interactions will include the mean square for treatments × replicates plus a component of variance due to high-order interactions. When only one replication is used, it is assumed that the component of variance due to high-order interactions is small, and that the high-order interactions mean square can be regarded as an estimate of error. It is important to bear in mind that an individual agricultural experiment can give information only for a particular set of experimental conditions and that it is known from experience that place to place and year to year variability is considerable. It would therefore be uneconomical to utilize available resources to determine effects and their errors at a few particular places very accurately, but preferable to sacrifice replication at each place in order to have information over a large range of experimental conditions.

Mixed systems

It is not proposed to examine mixed systems of the type $p^m q^n$, where p and q are primes, in the present paper. It is clear, however, that the possibilities of complete confounding and fractional replication are very limited. A p'th replicate must obviously include p^{m-1} combinations of the m factors combined with all the q^m combinations of the n factors. For the examination of treatment aliases the system may be regarded as the product of the two separate systems. Thus if $p = 3$, $m = 2$, $q = 2$, $n = 3$ and the factors are $y_1 y_2 y_3' y_4' y_5'$, then a half replicate would be obtained by putting $I = y_3' y_4' y_5'$. The aliases which result are exemplified by the following:

$$y_1 = y_1 y_3' y_4' y_5', \qquad y_1 y_2 y_3' = y_1 y_2 y_4' y_5',$$
$$y_1 y_2 = y_1 y_2 y_3' y_4' y_5', \qquad y_3' = y_4' y_5'.$$

Such designs with fractional replication or complete confounding are therefore useful only when the corresponding designs for the two separate systems are feasible.

Comments on 'The design of optimum multifactorial experiments'

In a paper entitled 'The Design of Optimum Multifactorial Experiments', Plackett & Burman (1946) put forward designs more specifically for physical and industrial research, which are of interest from the point of view of fractional replication. In order to estimate the

effect of varying nine components, of an assembly, each component having two possible values, a nominal ($-$) and an extreme ($+$), they put forward the following design which requires the testing of sixteen assemblies:

	Components 1	2	3	4	5	6	7	8	9
Assembly 1	+	−	−	−	+	−	−	+	+
2	+	+	−	−	−	+	−	−	+
3	+	+	+	−	−	−	+	−	−
4	+	+	+	+	−	−	−	+	−
5	−	+	+	+	+	−	−	−	+
6	+	−	+	+	+	+	−	−	−
7	−	+	−	+	+	+	+	−	−
8	+	−	+	−	+	+	+	+	−
9	+	+	−	+	−	+	+	+	+
10	−	+	+	−	+	−	+	+	+
11	−	−	+	+	−	+	−	+	+
12	+	−	−	+	+	−	+	−	+
13	−	+	−	−	+	+	−	+	−
14	−	−	+	−	−	+	+	−	+
15	−	−	−	+	−	−	+	+	−
16	−	−	−	−	−	−	−	−	−

Yates put forward a similar design in his 1935 paper for the weighing of a number of small articles on a balance which required a zero correction, as an example of the estimation of the effects of independent factors. In his case there was a close formal analogy to the 2^n factorial system, and it will now be shown that Plackett & Burman's design given above is a high-order fractional design of the type discussed in the present paper.

Denoting the nominal values by unity and the extreme values of the nine components by $a, b, c, d, e, f, g, h, k$ in order, the treatment combinations represented are l, $aehi$, $abfi$, $abcg$, $abcdh$, $bcdei$, $acdef$, $bdefg$, $acefgh$, $abdfghi$, $bceghi$, $cdfhi$, $adegi$, $befh$, $cfgi$, dgh. It is found merely by one-by-one examination of the three-factor interactions that all the above sets of treatment combinations occur with the same sign in the following:

$$ABE, \quad ACK, \quad BCF, \quad CDG, \quad DEH.$$

The same will be true for all the members of the Abelian group of which the above five interactions are generators. The identity relationship is therefore:

$$\begin{array}{llllllll}
I & = ABE & = ACK & = BCEK & = BCF & = ACEF & = ABFK & = EFK \\
= CDG & = ABCDEG & = ADGK & = BDEGK & = BDFG & = ADEFG & = ABCDFGK & = CDEFGK \\
= DEH & = ABDH & = ACDEHK & = BCDHK & = BCDEGH & = ACDFH & = ABDEFHK & = DFHK \\
= CEGH & = ABCGH & = AEGHK & = BGHK & = BEFGH & = AFGH & = ABCEFGHK & = CFGHK
\end{array}$$

The identities of interest to the experimenter are the following:

$$I = ABE = ACK = BCF = EFK = CDG = DEH;$$

from these we derive the following aliases for main effects:

$$\begin{array}{ll}
A = BE = CK, & F = BC = EK, \\
B = AE = CF, & G = CD, \\
C = AK = BF = DG, & H = DE, \\
D = CG = EH, & K = AC = EF. \\
E = AB = FK = DH, &
\end{array}$$

In all cases, the contrasts estimating main effects are minus the contrasts estimating interactions. If, for example, the interaction of B and E is negative, and A has no effect, the

conclusion drawn by the experimenter will be that A has a positive effect. It is possible but rather difficult to imagine physical systems in which effects will not interact, and interpretation of the results of experiments based on this design may often be impossible. With nine factors, it appears from the present work that the minimum number of combinations which should be tested is 128, that is one-quarter of a replication, though it is possible that by making less stringent assumptions about two-factor interactions, one-eighth of a replication might give intelligible results. A possible instance in which it might be feasible to use the designs discussed is when it is expected that only one or two of the factors have an effect, and the problem is to determine as quickly as possible which of the nine factors are responsible. An example in which a high-order fractional design was used in such circumstances with good results has been described by Tippett (1936). A detailed examination of all the designs put forward by Plackett & Burman will not be undertaken, but the lines on which such an examination would proceed and the broad conclusions which would emerge are obvious from the above examination of one of their simpler designs.

Conclusions

A method of examining fractional replication and confounding for some types of factorial experiments is described. The formal equivalence between the two is indicated and the implications of this equivalence discussed. Further progress will follow on group theory lines and this is being examined, together with the possibility of fractional replication when the fraction is greater than unity. The possibilities are explored of the estimation of main effects and two-factor interactions of many factors by testing only a small proportion of the possible treatment combinations. An examination on these lines is made of designs proposed by Plackett & Burman.

References

Finney, D. J. (1945). The fractional replication of factorial arrangements. *Ann. Eugen., Lond.*, **12**, 291–301.

Fisher, R. A. (1942). The theory of confounding in factorial experiments in relation to the theory of groups. *Ann. Eugen., Lond.*, **11**, 341–53.

Kempthorne, O. (1947). A note on differential responses in blocks. *J. Agric. Sci.* **37**, 245–48.

Plackett, R. L. & Burman, J. P. (1946). The design of optimum multifactorial experiments. *Biometrika*, **33**, 305–25.

Tippett, L. H. C. (1936). *Applications of Statistical Methods to the Control of Quality in Industrial Production.* Manchester Statist. Soc.

Yates, F. (1935). Complex experiments. *J. R. Statist. Soc. Suppl.* **2**, 181–223.

Yates, F. (1937). The design and analysis of factorial experiments. *Tech. Commun. Imp. Bur. Soil Sci.*, no. 35.

A COMPARISON OF STRATIFIED WITH UNRESTRICTED RANDOM SAMPLING FROM A FINITE POPULATION*

By P. ARMITAGE, B.A.

1. Introduction

1·1. We are concerned in this paper with the problem of estimating the mean value μ of a variable x in a population, by taking a sample which is in some way representative of the population. It has been realized since Bowley's paper (1926), and more particularly since Neyman's more comprehensive survey (1934), that a certain degree of precision in the estimate can often be obtained more economically by stratified random sampling (usually referred to merely as *stratified sampling*) than by unrestricted random sampling (usually called merely *random sampling*). In the stratified method, the population is divided into several strata, the sample size divided in some prearranged way among the strata, and sampling performed at random from each stratum. In unrestricted random sampling, a random selection is made from the whole population, and the method may be regarded as a particular case of stratification, where the number of strata is one.

Some text-books deal briefly with stratified sampling. Wilks (1943) considers only infinite populations, and denotes by representative sampling what we should call a particular type of stratified sampling (see § 1·2). The subject is treated by Kendall (1946, pp. 249–52), but he makes no comparison with unrestricted random sampling. We shall begin by introducing several well-known results which will be needed later.

1·2. The summation sign $\sum$ will be used throughout for $\sum\limits_{l=1}^{r}$, and $\sum\limits_{k}$ for $\sum\limits_{k=1}^{r}$. In general, $\sum$ is used for a single summation, $\sum\limits_{k}\sum$ for a double summation, and the suffix k where no summation is involved.

We shall consider the following position: A population π of size N is subdivided into r strata, π_k, of size N_k ($\sum N_l = N$). The variable x is distributed so that the mean and variance (divisor N_k) within π_k are respectively μ_k, σ_k^2. It is required to estimate $\mu = \sum N_l\mu_l/N$, the grand mean.

Suppose a given sample size, n, is divided so that n_k items are sampled at random from π_k ($\sum n_l = n$). We may denote the jth observation from the kth sample by x_{kj} ($j = 1, 2, \ldots, n_k$), and the mean and variance of the kth sample by $\bar{x}_k$ and s_k^2, which are known to be unbiased estimates of μ_k and $\dfrac{(n_k-1)\,N_k\sigma_k^2}{n_k(N_k-1)}$, respectively (see, for example, Kendall, 1943, p. 284). It seems intuitively obvious to take as our estimate of μ,

$$m = \sum N_l\bar{x}_l/N, \tag{i}$$

which is clearly unbiased. This is, however, not the only unbiased estimate which is a linear function of the x_{kj}. For instance, $\sum N_l x_{l1}/N$ also satisfies the conditions. Neyman (1934) has shown that, for fixed values of n_k, the estimate given by (i) is the best linear unbiased estimate of μ, in the sense that its sampling variance is less than that of any other linear unbiased estimate.

* Communication from the National Physical Laboratory.

The question now arises: given a sample size n, how shall we choose the n_k so as to minimize var (m), where m is given by (i)? Bowley had not considered 'best' estimates, and he suggested that n_k should be proportional to N_k, i.e.

$$n_k = \frac{nN_k}{N}. \tag{ii}$$

Neyman (1934) showed, by the method given in § 2, that the values of n_k which minimize var (m) are

$$\begin{aligned} n_k &= \frac{nN_k\sigma_k\sqrt{[N_k/(N_k-1)]}}{\Sigma N_l\sigma_l\sqrt{[N_l/(N_l-1)]}} \\ &= \frac{nN_k\sigma'_k}{\Sigma N_l\sigma'_l}, \end{aligned} \tag{iii}$$

where $\sigma'_k = \sigma_k\sqrt{[N_k/(N_k-1)]}$.

We shall refer to these two methods of defining the n_k, by (ii) and (iii) respectively, as *proportionate sampling*, and *optimum stratified sampling*, denoting by m_p and m_o the estimates of μ obtained from (i) by the two methods, and by $\bar{x}$ the estimate of μ given by the mean of an unrestricted random sample of n from the whole population π.

The optimum stratified method thus requires a knowledge of the σ_k. In practice, we should never know the σ_k exactly, unless the population had been subjected to exhaustive sampling, in which case μ would be known exactly. Sukhatme (1935) has shown that, at any rate for large N_k, if the σ_k^2 are estimated from a preliminary sample, and the n_k defined by using these estimates in (iii), there is a high probability that var (m_o) < var (m_p).* The efficiency of this method will of course depend on the size of the preliminary sample, and Sukhatme's investigation only dealt with one value of this (15 from each stratum). In some cases we should be able to form a fairly good estimate of the σ_k from past experience, and there would be no need for a preliminary sample.

Another interesting comparison which has not been extensively investigated is that between optimum stratified sampling and unrestricted random sampling. Wilks (1943) deals with this for infinite populations, and obtains (pp. 88, 89) the result (in our notation),

$$\text{var}(m_o) \leqslant \text{var}(m_p) \leqslant \text{var}(\bar{x}), \tag{iv}$$

the first equality holding only when all the σ_k are equal, and the second only when all the μ_k are equal. (Our N_k/N are replaced by p_k, where p_k is the probability that x, when drawn at random from π, is a member of π_k, so that, for instance, (iii) becomes

$$n_k = \frac{np_k\sigma_k}{\Sigma p_l\sigma_l}.$$

Representative sampling as defined by Wilks is what we should call proportionate sampling.) We shall show in § 2 that for finite populations, while the relation

$$\text{var}(m_o) \leqslant \text{var}(m_p) \tag{v}$$

is always true, the equality holding only when all the σ'_k are equal, it is not necessarily true that

$$\text{var}(m_p) \leqslant \text{var}(\bar{x}), \tag{vi}$$

* No confusion need arise from the fact that the symbol m_o and the term *optimum* are still used when estimates of the σ_k are used in (iii).

and in fact in the limiting case when all the μ_k are equal, it is true that

$$\operatorname{var}(m_p) > \operatorname{var}(\bar{x}), \tag{vi$a$}$$

so that if the σ'_k are also equal

$$\operatorname{var}(m_o) > \operatorname{var}(\bar{x}); \tag{vi$b$}$$

i.e. random sampling gives a more accurate estimate of the mean than any stratified sampling. We shall see, however, that in almost all practical cases (iv) is true.

2. Derivation of formulae

2·1. *Results* (iii) *and* (v). Using the notation of § 1·2, we have the standard result that

$$\operatorname{var}(\bar{x}_k) = \frac{\sigma_k^2}{n_k}\left(\frac{N_k - n_k}{N_k - 1}\right) \quad \text{(see e.g. Wilks, p. 86).} \tag{vii}$$

Therefore from (i),

$$\begin{aligned}\operatorname{var}(m) &= \sum \frac{N_l^2\sigma_l^2}{N^2 n_l}\left(\frac{N_l - n_l}{N_l - 1}\right) \\ &= \sum \frac{N_l \sigma_l'^2}{N^2 n_l}(N_l - n_l).\end{aligned} \tag{viii}$$

The result (iii) may be obtained quite easily by finding the values of the n_l which minimize (viii) subject to the condition $\sum n_l = n$, using the method of Lagrange multipliers. Then, substituting (ii) and (iii) in (viii), and applying Schwarz's inequality, we have (v). The following method is due to Neyman.

It may be verified from (viii) that

$$\operatorname{var}(m) = \frac{N-n}{N^2 n}\sum N_l\sigma_l'^2 + \frac{1}{N^2}\sum_k n_k\left(\frac{N_k\sigma'_k}{n_k} - \frac{\sum N_l\sigma'_l}{n}\right)^2 - \frac{1}{Nn}\sum_k N_k\left(\sigma'_k - \frac{\sum N_l\sigma'_l}{N}\right)^2. \tag{ix}$$

If we denote the three terms of (ix) by A, B and C, so that

$$\operatorname{var}(m) = A + B - C,$$

it will be seen that A and C are independent of n_k and, since B is non-negative, it follows that the values of n_k which minimize $\operatorname{var}(m)$ must minimize B. Now $B = 0$ if and only if

$$n_k = \frac{nN_k\sigma'_k}{\sum N_l\sigma'_l},$$

which is (iii). For these values of n_k, $m = m_o$, and

$$\operatorname{var}(m_o) = A - C. \tag{x}$$

If we define n_k by (ii), so that $m = m_p$, we see from (ix) that $B = C$, so that

$$\operatorname{var}(m_p) = A. \tag{xi}$$

From (x) and (xi), we obtain (v), the equality holding only when $C = 0$, which is true only when $\sigma'_k - \dfrac{\sum N_l\sigma'_l}{N} = 0$ for all k, i.e. when the σ'_k are all equal.

2·2. *Unrestricted random sampling.* The variance of a random observation x from π is

$$\sigma^2 = \frac{\sum N_l\sigma_l^2}{N} + S,$$

where S is the weighted sum of squares of the μ_l, i.e. $S = \dfrac{\sum N_l(\mu_l - \mu)^2}{N}$. From (vii),

$$\begin{aligned}\operatorname{var}(\bar{x}) &= \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)\\ &= \frac{N-n}{nN(N-1)}\sum N_l\sigma_l^2 + \frac{N-n}{n(N-1)}S\\ &= \frac{N-n}{nN(N-1)}\sum(N_l-1)\sigma_l'^2 + \frac{N-n}{n(N-1)}S.\end{aligned}$$

From (xi),
$$\operatorname{var}(m_p) = \frac{N-n}{N^2n}\sum N_l\sigma_l'^2.$$

Denoting $\dfrac{\sum(N_l-1)\sigma_l'^2}{N-1}$ by H, and $\dfrac{\sum N_l\sigma_l'^2}{N}$ by K, we have

$$\left.\begin{aligned}\operatorname{var}(\bar{x}) &= \frac{N-n}{Nn}H + \frac{N-n}{(N-1)n}S\\ \text{and}\qquad \operatorname{var}(m_p) &= \frac{N-n}{Nn}K.\end{aligned}\right\} \tag{xii}$$

Now
$$H - K = \frac{\sum N_l\sigma_l'^2 - \sum N_l \sum\sigma_l'^2}{N(N-1)} < 0,$$

and if we regard each N_l as being of the same order, $O(N)$, then $H-K$ is $O(N^{-1})$, which means that when all the μ_k are equal, $S = 0$, and so

$$\operatorname{var}(\bar{x}) < \operatorname{var}(m_p),$$

which is (vi$a$); but as $N \to \infty$,
$$\operatorname{var}(\bar{x}) \sim \operatorname{var}(m_p) + S/n, \tag{xiii}$$

giving Wilks's result (p. 88) that for infinite populations (vi) is true, the equality holding only when all the μ_k are equal.

From (v) and (vi$a$) it follows that for finite populations, when all the μ_k are equal and all the σ_k' are equal, (vi$b$) is true, i.e. in this case unrestricted random sampling is actually better than any stratified random sampling with the same sample size.

3. General comparison

3·1. From (ix), (x) and (xii),

$$\phi \equiv \operatorname{var}(\bar{x}) - \operatorname{var}(m_o) - \frac{N-n}{(N-1)n}S = \frac{1}{N^2(N-1)n}(P-Q-R), \tag{xiv}$$

where
$$\begin{aligned}P &= N^2\sum N_l\sigma_l'^2 - N(\sum N_l\sigma_l')^2 \geqslant 0 \quad \text{(equality if all } \sigma_l' \text{ are equal)},\\ Q &= n(\sum N_l\sigma_l'^2 - N\sum\sigma_l'^2) < 0,\\ R &= N^2\sum\sigma_l'^2 - (\sum N_l\sigma_l')^2 > 0.\end{aligned}$$

As $N \to \infty$, P, Q and R are respectively $O(N^3)$, $O(N)$ and $O(N^2)$, and so we have the result that for infinite populations $\phi \geqslant 0$, which with (xiii) is easily seen to be equivalent to Wilks's result (iv).

In the finite case, however, by suitable choice of the σ_l' and n we can make ϕ either positive or negative. For instance, if the σ_l' are all equal and n is sufficiently small, R predominates in (xiv), and $\phi < 0$. As n increases to N, ϕ increases to 0. (By considering Q and R, it is not

obvious that $\phi \to 0$ in this case, but it must be remembered that (xiv) is only true if the n_k are given by (iii), and this becomes impossible as n approaches N. This will be remarked upon below.) If the σ'_l are sufficiently unequal, P will predominate and $\phi > 0$. In this case the factor $\frac{(P-R)}{N^2(N-1)n}$ in (xiv) will be positive, and ϕ will decrease as n increases.

The situations, then, in which (vi$b$) is likely to be true (provided that the n_k are really given by (iii)) are when the μ_k are nearly equal, and when N is small or the σ'_k are nearly equal. We shall consider some examples in § 4.

3·2. In applying the procedure of stratification, we shall make two departures from the theory outlined above which will tend to nullify the advantages of the stratified method. The first is that, as was pointed out in § 1·2, we shall never know the σ_k exactly, and the degree to which our estimates from which the n_k were obtained are accurate depends on the circumstances. It seems quite likely that Sukhatme's result will be fairly well applicable to finite populations, but there is an opportunity for research on this point.

The second respect in which we depart from theory lies in the fact that, even if the σ_k are exactly known, the n_k that we choose can never be exactly as given by (iii); first because they must be integers, which makes a considerable difference when n is small (the size of the smallest stratified sample from which an unbiased estimate of μ can be made is clearly r); and secondly, n_k cannot take values greater than N_k. In this latter case, if the values of, say, s of the n_k, as given by (iii), are greater than the corresponding N_k, we should let $n_k = N_k$ for these s strata, and then set the other $(r-s)$ values of n_k proportional to the corresponding $N_k\sigma'_k$. This will clearly decrease $\text{var}(\bar{x}) - \text{var}(m_o)$ as given by (xiv). For example, when $n = N$, we have

$$\text{var}(\bar{x}) = \text{var}(m_o) = \frac{(N-n)}{(N-1)n} S = 0,$$

but the right-hand side of (xiv)

$$= \frac{1}{N^3}\{N\sum N_l\sigma_l'^2 - (\sum N_l\sigma'_l)^2\} \geqslant 0$$

(equality holding if all the σ'_l are equal). In fact both these limitations will decrease the theoretical advantage (if any) of stratified over random sampling, and we must take them into account in assessing the relative merits of the two methods.

4. EXAMPLES

In the four examples illustrated by Figs. 1–4, $\text{var}(m_o)$ and $\text{var}(\bar{x})$ have been calculated for different stratified populations, and $\psi = \log_{10}\{\text{var}(\bar{x})/\text{var}(m_o)\}$ plotted against $c = n/N$, so that $\psi < 0$ if $\text{var}(\bar{x}) < \text{var}(m_o)$. In each figure the different curves represent populations with the same σ_k, with the N_k in the same proportions but with different magnitudes, and with the μ_k equal, so that $S = 0$.

Example 1.	$\sigma_k = 2, 3, 4,$	$N_k \propto 6, 4, 3$	$(N = 65, 26, 13)$.
Example 2.	$\sigma_k = 4, 5, 6,$	$N_k \propto 6, 5, 4$	$(N = 120, 60, 30, 15)$.
Example 3.	$\sigma_k = 4, 5, 6,$	$N_k \propto 3, 11, 4$	$(N = 126, 54, 36, 18)$.
Example 4.	$\sigma_k = 1, 1, 2, 3, 4,$	$N_k \propto 5, 5, 1, 2, 3$	$(N = 128, 32)$.

The first thing to be noticed about the graphs is that in each one ψ increases, generally speaking, as n increases. Further, in any one example the range of c for which $\psi < 0$ increases

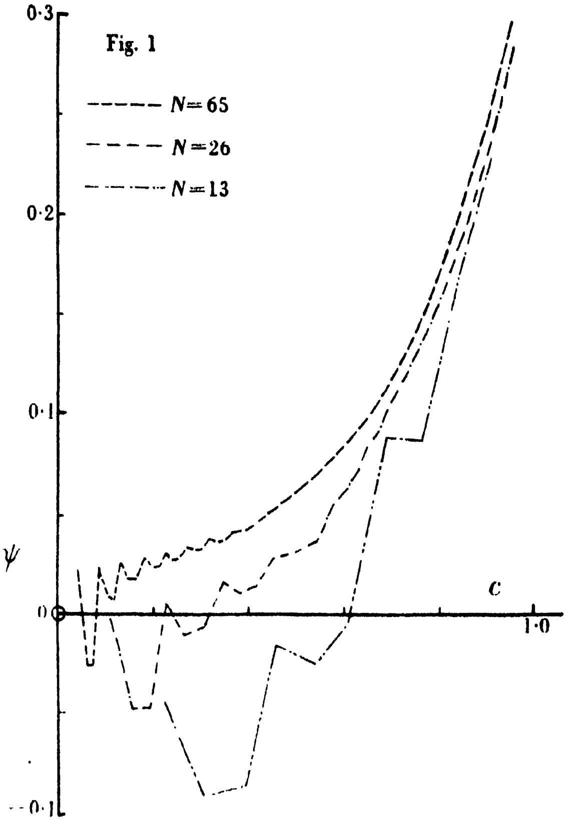


Fig. 1
N=65
N=26
N=13
0·3
0·2
0·1
ψ
0
−0·1
c
1·0

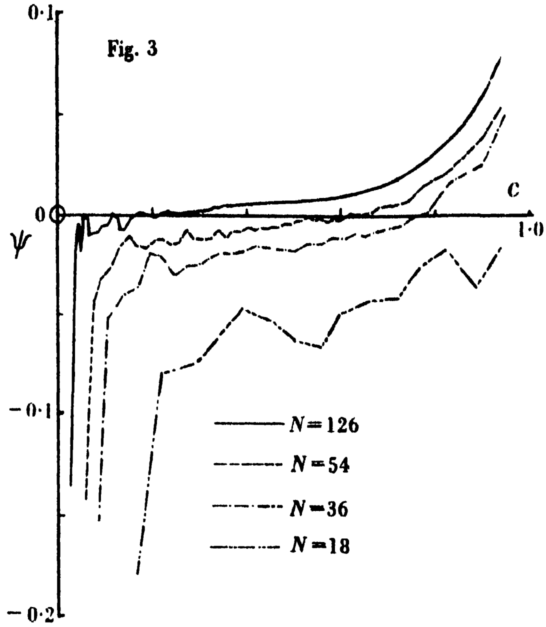


Fig. 3
0·1
0
ψ
−0·1
−0·2
c
1·0
N=126
N=54
N=36
N=18

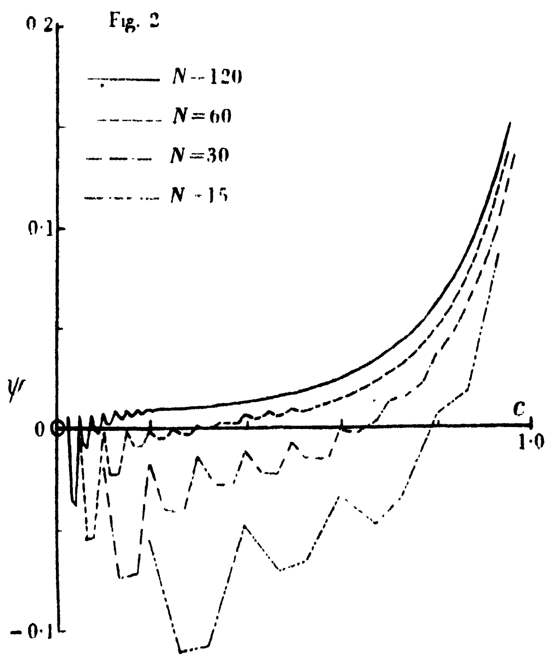


Fig. 2
N=120
N=60
N=30
N=15
0·2
0·1
ψ
0
−0·1
c
1·0

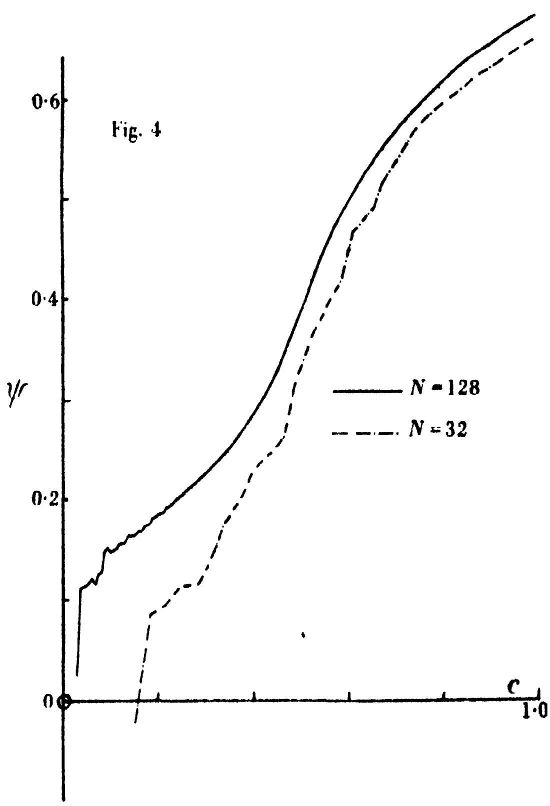


Fig. 4
0·6
0·4
ψ
0·2
0
N=128
N=32
c
1·0

as N decreases; and in this sense we can say that for small samples of proportionate size from a stratified population, the advantage (if any) of the stratified method decreases as N decreases.

Secondly, the curves are not smooth. The reason for this is clear. In the optimum stratified method the n_k are to be chosen approximately proportional to $N_k\sigma_k$ (a second approximation is $(N_k+\frac{1}{2})\sigma_k$). In Example 1, the $N_k\sigma_k$ are all equal, and it follows that the n_k should be nearly equal. If $n \equiv 0$ (mod 3) this can be done, but for $n \equiv 1, 2$ (mod 3), var (m_o) takes values greater than it would if fractional n_k were allowed. This produces a rise in the curve of ψ for $n \equiv 0$ (mod 3), which gradually disappears as n increases since the effect is much greater for small n. The same 'period' is noticeable in Fig. 2, but in Figs. 3 and 4, where the main 'periods' are respectively 15 and 30, the effect is smaller.

We saw in § 3·1 that, broadly speaking, the advantage of the stratified method decreases as the σ_k tend to equality. This is illustrated by comparing Examples 1 and 2. In each of these the $N_k\sigma_k$ are equal, but in Example 2 the σ_k are proportionally more nearly equal. Comparing curves for about the same N ($N = 65, 26, 13$ in Fig. 1 with $N = 60, 30, 15$ in Fig. 2), we see that in Fig. 2 the range of values of c for which $\psi < 0$ is greater than in Example 1.

Fig. 3 has the same σ_k as Fig. 2, but the $N_k\sigma_k$, and therefore the n_k, are different. The curves are similar to those of Fig. 2, but the stratified method is still less advantageous (especially for small values of c).

Example 4 has five instead of three strata, and there is quite large variation between the σ_k and between the $N_k\sigma_k$. There is no doubt here that $\psi > 0$, the only exception being for $N = 32$, $n = 5$, where $\psi = -0{\cdot}02$.

These examples may be said to give the maximum advantage to the stratified method, in the sense that the calculated values of var (m_o) depend on the best method of choosing the n_k. If the σ_k are not sufficiently well known to enable the best values of n_k to be used, then we shall get a larger value of var (m_o). It must be remembered, however, that in all these examples we assumed that there was no variation between the μ_k, a situation which would be very unlikely to occur in practice. Now it is clear from (xii) that if the same N_k and σ_k are considered as in one of the above examples, but the μ_k are now unequal, the effect is to increase the value of var $(\bar{x})$ by $(N-n)S/(N-1)n$, where $S = \sum N_l(\mu_l-\mu)^2/N$; so, in any example where $\psi < 0$ for some particular values of N and n, we can reverse the direction of the inequality by choosing a sufficiently large value of S, say

$$S_0 = [\operatorname{var}(m_o) - \operatorname{var}(\bar{x})](N-1)n/(N-n).$$

In comparing different values of S_0 for different examples, it must be remembered that the order of magnitude of S_0 depends on the σ_k and a suitable measure of comparison will be S_0/σ_0^2, where σ_0^2 is the pooled variance within strata $= \sum N_l\sigma_l^2/N$.

In Example 1, the largest value of S_0 is for $N = 13$, $n = 4$. Here var $(m_o) = 1{\cdot}9172$, var $(\bar{x}) = 1{\cdot}5577$, and $S_0 = 1{\cdot}917 = 0{\cdot}231\sigma_0^2$. (If $\mu_1 = 0$, $\mu_2 = 2$, $\mu_3 = 3{\cdot}5$, then $S = 2{\cdot}066$.)

In Example 2, the largest value of S_0 is for $N = 15$, $n = 4$. Here var $(m_o) = 24{\cdot}647$, var $(\bar{x}) = 19{\cdot}119$, and $S_0 = 28{\cdot}14 = 0{\cdot}289\sigma_0^2$. (If $\mu_1 = 0$, $\mu_2 = 7$, $\mu_3 = 13$, then $S = 28{\cdot}51$.)

In Example 3, the largest value of S_0 is for $N = 18$, $n = 3$. Here var $(m_o) = 46{\cdot}235$, var $(\bar{x}) = 30{\cdot}523$, and $S_0 = 53{\cdot}42 = 0{\cdot}515\sigma_0^2$. (If $\mu_1 = 0$, $\mu_2 = 8$, $\mu_3 = 17$, then $S = 57{\cdot}5$.)

In Example 4, the largest value of S_0 is for $N = 32$, $n = 5$ (the only occasion in this example where $\psi < 0$). Here var $(m_o) = 0{\cdot}91406$, var $(\bar{x}) = 0{\cdot}87097$, and $S_0 = 0{\cdot}2474 = 0{\cdot}049\sigma_0^2$. (If $\mu_1 = \mu_2 = 0$ and $\mu_3 = \mu_4 = \mu_5 = 1$, then $S = 0{\cdot}285$.)

5. Conclusions

We have seen in § 3 that optimum stratified sampling may give a less accurate estimate of μ than unrestricted random sampling when the μ_k are nearly equal, and when N is small or the σ'_k are nearly equal. The examples of § 4 bear out these conclusions and show that the effect is greatest for small n, Fig. 3 providing an additional suggestion that if the products $N_k\sigma_k$ are widely different the advantage of the stratified method tends to be nullified. In practice, we should probably only apply stratified sampling if we knew that the strata were sufficiently distinct to ensure considerable variation between either the μ_k or the σ_k. In the first case, if nothing much was known about the σ_k and a preliminary sample on the lines suggested by Sukhatme was impracticable, we should use proportionate sampling, and the size of S would usually ensure that $\operatorname{var}(m_p) < \operatorname{var}(\bar{x})$. In the second case, we should use optimum stratified sampling, and rely on the variability of the σ_k to ensure that $\operatorname{var}(m_o) < \operatorname{var}(\bar{x})$. Since an adequate degree of knowledge about the σ_k would be unlikely unless the N_k were quite large, we should in this case almost certainly be safe in using the method. To the above considerations must be added the fact that if very inaccurate estimates of the σ_k are used in (iii), then, whatever the nature of the population, the resulting procedure may be extremely inefficient.

It must be realized, of course, that even if it were known that $\operatorname{var}(m_o) < \operatorname{var}(\bar{x})$, it would not follow that the optimum stratified method would necessarily be the most convenient. It may be impossible, or at any rate inconvenient, to do any sort of random sampling, and some sort of quasi-random sampling may have to be used (see. e.g. Madow & Madow, 1944), but if the principle of random sampling is applicable the stratified method is not likely to be much more inconvenient, and in fact in most cases will be more convenient, than the unrestricted method.

Summary

The stratified method has been used in the past almost solely for large-scale social and agricultural surveys. Here the stratum sizes are large, and known results for infinite populations apply. There seems no reason why stratified sampling should not be used to advantage for smaller populations, and it is important to know to what extent these results still apply. In this paper a comparison has been made with unrestricted random sampling in the usual case where we are interested in estimating the mean. The advantages of the stratified method are modified, but in most cases where the method is applicable it will be found to be worth while.

The above work was carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory. The author desires to acknowledge the assistance rendered by Mr D. V. Lindley who prepared the diagrams.

REFERENCES

Bowley, A. L. (1926). Measurement of the precision attained in sampling. *Bull. Int. Inst. Statist.* **22**, 1ère livraison.

Kendall, M. G. (1943). *The Advanced Theory of Statistics*, **1**. Griffin and Co.

Kendall, M. G. (1946). *The Advanced Theory of Statistics*, **2**. Griffin and Co.

Madow, W. G. & Madow, L. H. (1944). On the theory of systematic sampling. *Ann. Math. Statist.* **15**, 1–24.

Neyman, J. (1934). On the two different aspects of the representative method. *J. Roy. Statist. Soc.* **91**, 558–625.

Sukhatme, P. V. (1935). Contribution to the theory of the representative method. *Suppl. J. Roy. Statist. Soc.* **2**, 253–68.

Wilks, S. S. (1943). *Mathematical Statistics*. Princeton University Press.

SOME THEOREMS ON TIME SERIES. I

By P. A. P. MORAN
Institute of Statistics, Oxford University

One of the principal problems in the theory of time series is to discuss the relation between two series, and in the present paper we prove a theorem by which we can test whether two such series are independent. Such a test of significance must depend on the models which we assume for the probability processes which generate the series. In practice, the two most useful models are, first, that of a moving average of a series of independent random components and, secondly, the solutions of linear stochastic difference equations.

Let
$$\ldots, \eta(t-1), \eta(t), \eta(t+1), \ldots$$
be a sequence of independent random variables each distributed in the same distribution which we take to have zero mean and its second, third, and fourth moments finite. Then the time series generated by
$$X(t) = \sum_{i=0}^{N} \alpha_i \eta(t-i)$$
is a moving average with weights α_i. On the other hand, consider a stochastic difference equation of the form
$$X(t) + a_1 X(t-1) + \ldots + a_h X(t-h) = \eta(t). \tag{1}$$

In order that the solution of (1) for successive values of t shall form a stationary series it is necessary to impose the condition that the roots of the characteristic equation
$$z^h + a_1 z^{h-1} + \ldots + a_h = 0 \tag{2}$$
shall all lie inside the circle $|z| = 1$ (Wold, 1938, p. 53). When this is true the solution of (1) can be shown to be of the form
$$X(t) = \sum_{i=0}^{\infty} \alpha_i \eta(t-i),$$
where the α_i are certain functions of the roots of (2). In this case $\sum_{i=0}^{\infty} |\alpha_i|$ is majorized by a convergent geometric series.

Thus we see that both the above models are included in the more general one in which we define $X(t)$ as given by
$$X(t) = \sum_{i=0}^{\infty} \alpha_i \eta(t-i),$$
where the α_i are any sequence of constants satisfying $\sum_{i=0}^{\infty} |\alpha_i| < \infty$. Now suppose
$$\ldots, \zeta(t-1), \zeta(t), \zeta(t+1), \ldots$$
is another sequence of independent random variables having a distribution with zero mean and finite second, third and fourth moments. We write
$$Y(t) = \sum_{i=0}^{\infty} \beta_i \zeta(t-i),$$
where $\sum_{i=0}^{\infty} |\beta_i| < \infty$. To discuss whether two such empirical series of this form are correlated we prove that the covariance
$$S = \sum_{1}^{n} X(t)\, Y(t) \tag{3}$$

tends, as n increases, to be distributed in the normal form about zero mean with a second moment which is a function of the α_i and the β_i. We shall discuss later the calculation of this second moment from empirical series, in which case some care is necessary.

We first illustrate our method of proof by considering the much simpler problem of determining the asymptotic distribution of the sum

$$T_n = \sum_{s=1}^{n} X(t-s). \tag{4}$$

We shall show that this asymptotic distribution is also, under certain conditions, normal. This result is interesting because it establishes a central limit theorem (and therefore a law of large numbers) for stationary stochastic processes of this type. The law of large numbers for Markov chains has been considered by several writers, in particular Bernstein (1927), who proves his results by using central limit theorems for non-independent components. His theorems cannot be applied in the present case, but some of the ideas of his methods can.

Consider (4) above, where $X(t)$ is defined by

$$X(t) = \sum_{i=0}^{\infty} \alpha_i \eta(t-i)$$

and $\sum_{i=0}^{\infty} |\alpha_i|$ is convergent. There is no loss in generality in supposing that

$$\sum_{i=0}^{\infty} |\alpha_i| < 1.$$

Clearly
$$E(T_n) = \sum_{s=1}^{n} \sum_{i=0}^{\infty} \alpha_i E[\eta(t-s-i)]$$
$$= 0.$$

Write
$$\sigma^2 = E(\eta^2), \quad c_0 = E[X(t)^2], \quad c_s = E[X(t)X(t-s)].$$

Then
$$c_0 = \sigma^2(\alpha_0^2+\alpha_1^2+\dots), \quad c_s = \sigma^2(\alpha_0\alpha_s+\alpha_1\alpha_{s+1}+\dots),$$

which are both clearly convergent. Moreover,

$$R_n = E(T_n^2) = nE[X(t)^2] + 2\sum_{i=1}^{n-1}\sum_{s=1}^{n-i} E[X(t-i)X(t-i-s)]$$
$$= \left(nc_0 + 2\sum_{i=1}^{n-1}(n-i)c_i\right).$$

$n^{-1}R_n$ tends, as n increases, to
$$R_0 = \left(c_0 + 2\sum_{i=1}^{\infty} c_i\right)$$

if this series converges absolutely. We shall show that $\lim n^{-1}R_n$ is finite. For R_0 is clearly

$$\left(\sum_{i=0}^{\infty}\alpha_i\right)^2\sigma^2, \tag{5}$$

and this is finite. Moreover, we notice that $n^{-1}R_n$ is not greater than

$$\left(\sum_{i=0}^{\infty}|\alpha_i|\right)^2\sigma^2.$$

We must now impose the condition that $\sum_{i=0}^{\infty}\alpha_i$ is not zero. This condition is necessary to our

method of argument. If it is not zero, it may be assumed, without loss of generality, greater than a positive number. We now show that as n increases

$$pr\{t_0(2R_n)^{\frac{1}{2}} \leqslant T_n < t_1(2R_n)^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to $$\pi^{-\frac{1}{2}}\int_{t_0}^{t_1} e^{-t^2}\,dt.$$

We require the following lemma (Bernstein, 1927, p. 12):

LEMMA I. Let $$\rho_n = \Sigma_n + \sigma_n,$$
where ρ_n, Σ_n and σ_n are random variables such that

$$E(\Sigma_n) = E(\sigma_n) = 0, \quad E(\Sigma_n^2) = H_n, \quad E(\sigma_n^2) = H'_n.$$

Then if, for n large, $$pr\{t_0(2H_n)^{\frac{1}{2}} \leqslant \Sigma_n < t_1(2H_n)^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to $$\pi^{-\frac{1}{2}}\int_{t_0}^{t_1} e^{-t^2}\,dt,$$

then $$pr\{t_0(2J_n)^{\frac{1}{2}} \leqslant \rho_n < t_1(2J_n)^{\frac{1}{2}}\},$$
where $J_n \doteq E(\rho_n^2)$ tends, uniformly in t_0 and t_1, to

$$\pi^{-\frac{1}{2}}\int_{t_0}^{t_1} e^{-t^2}\,dt,$$

provided that $$\lim_{n\to\infty} \frac{H'_n}{H_n} = 0.$$

Let ϵ be an arbitrarily small number and choose N so large that

$$\sum_{i=N}^{\infty} |\alpha_i| < \epsilon \sum_{i=0}^{\infty} |\alpha_i| < \epsilon.$$

Write $$X_1(t) = \sum_{i=0}^{N} \alpha_i \eta(t-i), \quad T'_n = \sum_{s=1}^{n} X_1(t-s).$$

Then $$E(T'_n) = 0,$$

and write $$R'_n = E(T'^2_n).$$

We shall prove that the distribution of T'_n tends to normality, i.e. that

$$pr\{t_0(2R'_n)^{\frac{1}{2}} \leqslant T'_n < t_1(2R'_n)^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to $$\pi^{-\frac{1}{2}}\int_{t_0}^{t_1} e^{-t^2}\,dt.$$

We first calculate R_n and R'_n in another way. For

$$\begin{aligned} T_n &= \sum_{s=1}^{n} X(t-s) = \sum_{s=1}^{n} \sum_{i=0}^{\infty} \alpha_i \eta(t-s-i) \\ &= \alpha_0 \eta(t-1) + (\alpha_0+\alpha_1)\,\eta(t-2) + \ldots + (\alpha_0 + \ldots + \alpha_{n-1})\,\eta(t-n) \\ &\quad + \sum_{s=1}^{\infty} (\alpha_s + \ldots + \alpha_{s+n-1})\,\eta(t-s-n), \end{aligned}$$

and so $R_n = E(T_n^2) = \sigma^2 \left\{\alpha_0^2 + (\alpha_0+\alpha_1)^2 + \ldots + (\alpha_0 + \ldots + \alpha_{n-1})^2 + \sum_{s=1}^{\infty} (\alpha_s + \ldots + \alpha_{s+n-1})^2\right\},$

and this series converges. On the other hand,

$$T'_n = \alpha_0\eta(t-1)+(\alpha_0+\alpha_1)\eta(t-2)+\ldots+(\alpha_0+\ldots+\alpha_{N-1})\eta(t-N)$$
$$+\sum_{p=1}^{n-N}(\alpha_0+\ldots+\alpha_N)\eta(t-N-p)$$
$$+(\alpha_1+\ldots+\alpha_N)\eta(t-n-1)+\ldots+\alpha_N\eta(t-n-N),$$

and so
$$R'^2_n = \sigma^2\{\alpha_0^2+\ldots(\alpha_0+\alpha_1)^2+\ldots+(\alpha_0+\ldots+\alpha_{N-1})^2$$
$$+(n-N)(\alpha_0+\ldots+\alpha_N)^2+(\alpha_1+\ldots+\alpha_N)^2+\ldots+\alpha_N^2\}.$$

Since we have already supposed that $\sum_{i=0}^{\infty}\alpha_i$ is positive, there exist positive numbers N_0 and d such that for all $N>N_0$, $\sum_{i=0}^{N}\alpha_i>d$. If this is not true the theorem is in general false. For suppose the distribution of the η's to be non-normal and write $\alpha_0=1$, $\alpha_1=-1$, $\alpha_i=0$ $(i>1)$. Then the distribution of T_n does not tend to normality and its variance does not increase with n. We shall later show that this condition on the α's is in fact satisfied for the solutions of stochastic difference equations.

Now by the ordinary central limit theorem, as n increases,

$$T''_n = \sum_{p=1}^{n-N}(\alpha_0+\ldots+\alpha_N)\eta(t-N-p)$$

tends to be distributed normally with zero mean and variance

$$R''_n = (n-N)(\alpha_0+\ldots+\alpha_N)^2\sigma^2,$$

that is
$$pr\{t_0(2R''_n)^{\frac{1}{2}}\leqslant T''_n<t_1(2R''_n)^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to
$$\pi^{-\frac{1}{2}}\int_{t_0}^{t_1}e^{-t^2}\,dt.$$

Using Lemma I we see that the same is true, for fixed N, when we replace T''_n by T'_n and R''_n by R'_n. Now

$$T_n = T'_n+Q,$$

say, where Q is what we get if we replace the sequence $(\alpha_0,\alpha_1,\ldots)$ in T_n by $(\alpha_{N+1},\ldots)$ and alter t, and from (5) we can choose N so large that for $n>N$, $n^{-1}E(Q^2)<\epsilon$, say. Taking a sequence ϵ_1, ϵ_2, ... tending to zero and choosing first N sufficiently large and then n and using Lemma I again, we see that

$$pr\{t_0(2R_n)^{\frac{1}{2}}\leqslant T_n<t_1(2R_n)^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to
$$\pi^{-\frac{1}{2}}\int_{t_0}^{t_1}e^{-t^2}\,dt.$$

To complete the discussion we must show that the condition we have imposed on the sequence $\alpha_0,\alpha_1,\ldots$ is satisfied by the coefficients of the solutions of stationary stochastic difference equations. Consider an equation

$$X(t)+a_1X(t-1)+\ldots+a_hX(t-h) = \eta(t)$$

such that the roots of
$$z^h+a_1z^{h-1}+\ldots+a_h = 0 \tag{6}$$

all lie inside the circle $|z|=1$. Then the solution of this equation is given (Wold, 1938, p. 53) by

$$X(t) = \sum_{i=0}^{\infty}\alpha_i\eta(t-i),$$

where the α_i are now the solutions of the infinite set of equations

$$\begin{aligned}
&\alpha_0 = 1 \\
&a_1\alpha_0 + \alpha_1 &&= 0, \\
&a_2\alpha_0 + a_1\alpha_1 \quad + \alpha_2 &&= 0, \\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
&a_h\alpha_0 + a_{h-1}\alpha_1 + \dots + \alpha_h &&= 0, \\
&\qquad a_h\alpha_1 \quad + \dots + a_1\alpha_h + \alpha_{h+1} &&= 0, \\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots
\end{aligned}$$

and since the left-hand side is an absolutely convergent double series, we add, obtaining

$$(1 + a_1 + \dots + a_h) \sum_{i=0}^{\infty} \alpha_i = 1,$$

and so $\sum_{i=0}^{\infty} \alpha_i \neq 0$ and, as already observed, without loss of generality, may be supposed positive. This quantity is finite because all the roots of equation (6) lie inside the circle $|z| = 1$. Moreover, it follows that

$$R_0 = \left(\sum_{i=0}^{\infty} \alpha_i\right)^2 \sigma^2 = (1 + a_1 + \dots + a_n)^{-2}\sigma^2.$$

This is, in fact, proportional to the derivative at zero of the integrated power spectrum (Wold, 1938, p. 69).

We now turn to the problem of discussing the relation between two such series and we consider the asymptotic distribution of

$$S_n = \sum_{t=1}^{n} X(-t)\, Y(-t),$$

where

$$X(t) = \sum_{i=1}^{\infty} \alpha_i \eta(t-i) \quad (\alpha_1 \neq 0), \tag{7}$$

and

$$Y(t) = \sum_{i=1}^{\infty} \beta_i \zeta(t-i) \quad (\beta_1 \neq 0). \tag{8}$$

We write S_n in this form rather than that of (3) for the sake of convenience in what follows, and we have altered the notation of the sums (7) and (8) so that they begin with the coefficients α_1 and β_1 for the same reason. Writing

$$c_s = E[X(t)\,X(t-s)], \quad d_s = E[Y(t)\,Y(t-s)] \quad (s = 0, 1, \dots),$$

as before, we have

$$c_s = \sigma_1^2(\alpha_1\alpha_{s+1} + \alpha_2\alpha_{s+2} + \dots), \quad d_s = \sigma_2^2(\beta_1\beta_{s+1} + \beta_2\beta_{s+2} + \dots),$$

where σ_1^2 and σ_2^2 are the second moments of η and ζ. Then

$$E(S_n) = \sum_{t=1}^{n} E(X(-t)\,Y(-t)) = 0,$$

$$\begin{aligned}
E(S_n^2) &= E\left\{\sum_{t=1}^{n} X(-t)\,Y(-t)\right\}^2 \\
&= E\left\{\sum_{t=1}^{n} X^2(-t)\,Y^2(-t) + 2\sum_{t=1}^{n-1}\sum_{s=1}^{n-t} X(-t)\,X(-t-s)\,Y(-t)\,Y(-t-s)\right\} \\
&= nc_0d_0 + 2\sum_{s=1}^{n-1}(n-s)\,c_s d_s.
\end{aligned} \tag{9}$$

Consider the behaviour of $n^{-1}E(S_n^2)$ as n increases. Clearly

$$n^{-1}E(S_n^2) \to c_0 d_0 + 2 \sum_{s=1}^{\infty} c_s d_s = C, \quad \text{say}, \tag{10}$$

if the series C is absolutely convergent. If X and Y are moving averages or the solutions of stationary stochastic difference equations this is certainly true, for in the first case the series is finite, and in the second it is majorized by a convergent geometric series. We show that it is true in the general case by the following argument. Without restricting generality, we may assume, as before, that $\sum_1^{\infty} |\alpha_i| < 1$, $\sum_1^{\infty} |\beta_i| < 1$. Then

$$\begin{aligned} |c_s| &\leqslant \sigma_1^2(|\alpha_1 \alpha_s| + \ldots) \\ &\leqslant \sigma_1^2(|\alpha_1| + \ldots), \end{aligned}$$

and so

$$\begin{aligned} \left|\sum_1^{\infty} c_s d_s\right| &\leqslant \sum_1^{\infty} |c_s d_s| \leqslant \sigma_1^2 \sum_1^{\infty} |d_s| \\ &\leqslant \sigma_1^2 \sigma_2^2 \sum_{s=1}^{\infty} (|\beta_1|\,|\beta_s| + \ldots) \\ &\leqslant \sigma_1^2 \sigma_2^2 \left(\sum_{i=1}^{\infty} |\beta_i|\right)^2. \end{aligned} \tag{11}$$

Also

$$c_0 d_0 \leqslant \sigma_1^2 \sigma_2^2 \left(\sum_{i=1}^{\infty} |\alpha_i|^2\right)\left(\sum_{i=1}^{\infty} |\beta_i|^2\right),$$

and so $c_0 d_0 + 2 \sum_1^{\infty} c_s d_s$ is finite. We now prove that C is not zero. For

$$C = c_0 d_0 + 2 \sum_{s=1}^{\infty} c_s d_s = \sigma_1^2 \sigma_2^2 \left[(\alpha_1^2 + \alpha_2^2 + \ldots)(\beta_1^2 + \beta_2^2 + \ldots) + 2 \sum_{s=1}^{\infty} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \alpha_m \alpha_{m+s} \beta_n \beta_{n+s}\right],$$

and after some rearrangement, this equals

$$\sigma_1^2 \sigma_2^2 [(\alpha_1 \beta_1)^2 + (\alpha_1 \beta_2 + \alpha_2 \beta_1)^2 + (\alpha_1 \beta_3 + \alpha_2 \beta_2 + \alpha_3 \beta_1)^2 + \ldots],$$

and $(\alpha_1 \beta_1)^2$ is greater than zero and the rest non-negative at least. We therefore conclude that

$$n^{-1}E(S_n^2) \to C,$$

where

$$0 < C < \infty.$$

Assuming as before that
$$\sum_1^{\infty} |\alpha_i| < 1, \quad \sum_1^{\infty} |\beta_i| < 1,$$

we define N so that

$$\sum_{N+1}^{\infty} |\alpha_i| < \epsilon \sum_1^{\infty} |\alpha_i| < \epsilon, \tag{12}$$

$$\sum_{N+1}^{\infty} |\beta_i| < \epsilon \sum_1^{\infty} |\beta_i| < \epsilon, \quad \text{where } \epsilon \text{ is small.} \tag{13}$$

We now write

$$X_1(t) = \sum_{i=1}^{N} \alpha_i \eta(t-i), \tag{14}$$

$$Y_1(t) = \sum_{i=1}^{N} \beta_i \zeta(t-i), \tag{15}$$

and consider the sum

$$S_n' = \sum_1^{n} X_1(-t) Y_1(-t). \tag{16}$$

We begin by proving that when n is large this sum tends to be distributed in the normal form with a variance which is asymptotically equal to nC_1, where C_1 is obtained from C by putting $\alpha_i = \beta_i = 0$ for $i > N$. For it is then clearly true that

$$n^{-1}E(S_n'^2) \to C_1.$$

Now consider
$$S_n' = \sum_1^n X_1(-t)\,Y_1(-t),$$

where n is greater than N. For convenience of notation, we write

$$\eta_i = \eta(-i), \quad \zeta_i = \zeta(-i).$$

We then have
$$S_n' = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} A_{ij}\eta_i\zeta_j,$$

where the A_{ij} are certain constants. Moreover

$$\begin{aligned}
&E(\eta_i\zeta_j) = 0 && \text{all } i, j,\\
&E(\eta_i^2\zeta_j^2) = \sigma_1^2\sigma_2^2 && \text{all } i, j,\\
&E(\eta_i^2\zeta_j\zeta_k) = E(\eta_j\eta_k\zeta_i^2) = 0 && \text{for } j \neq k,\\
&E(\eta_i\eta_j\zeta_k\zeta_l) = 0 && \text{if } \; i \neq j \; \text{ or } \; k \neq l.
\end{aligned}$$

It therefore follows that
$$E(S_n'^2) = \sigma_1^2\sigma_2^2 \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} A_{ij}^2.$$

Inserting (14) and (15) in (16) we have

$$A_{ij} = 0$$

if
$$i > n+N, \quad \text{or} \quad j > n+N \quad \text{or} \quad |i-j| > N-1,$$

and
$$S_n' = \Sigma_1 + \Sigma_2 + \Sigma_3 + \Sigma_4,$$

where
$$\Sigma_1 = \sum_{i=1}^{N}\sum_{j=1}^{N} A_{ij}\eta_i\zeta_j$$

with
$$\begin{aligned}
A_{ij} &= \alpha_i\beta_j + \alpha_{i-1}\beta_{j-1} + \dots + \alpha_{1+i-j}\beta_1 && \text{for } \; i > j,\\
&= \alpha_i\beta_j + \qquad\qquad + \alpha_1\beta_{1+j-i} && \text{for } \; i < j,\\
&= \alpha_1\beta_1 + \qquad\qquad + \alpha_i\beta_j && \text{for } \; i = j.
\end{aligned}$$

We also have
$$\Sigma_2 = \Sigma\Sigma A_{ij}\eta_i\zeta_j,$$

where the sum is taken over values of i and j such that $|i-j| < N$, $i \leqslant n$, $j \leqslant n$ and either $N < i$ or $N < j$, where

$$\begin{aligned}
A_{ij} &= \alpha_1\beta_{p+1} + \dots + \alpha_{N-p}\beta_N && \text{for } \; j-i = p > 0\\
&= \alpha_{p+1}\beta_1 + \dots + \alpha_N\beta_{N-p} && \text{for } \; i-j = p > 0\\
&= \alpha_1\beta_1 \;\; + \dots + \alpha_N\beta_N && \text{for } \; i = j.
\end{aligned}$$

Then
$$E(\Sigma_2) = 0, \quad E(\Sigma_2^2) = \sigma_1^2\sigma_2^2\Sigma\Sigma A_{ij}^2,$$

where the sum is taken over the above values of i and j. This equals

$$\begin{aligned}
(n-N)\,\sigma_1^2\sigma_2^2[(\alpha_1\beta_N)^2 + (\alpha_1\beta_{N-1} + \alpha_2\beta_N)^2 + \dots\\
+ (\alpha_1\beta_1 + \dots + \alpha_N\beta_N)^2 + \dots + (\alpha_N\beta_2 + \alpha_{N-1}\beta_1)^2 + (\alpha_N\beta_1)^2].
\end{aligned} \tag{17}$$

We know that $\alpha_1 \neq 0$. Let β_t be the first term of the sequence $\beta_N, \beta_{N-1}, \ldots, \beta_1$ which is not zero. Such a term certainly exists. Then the sum in the outer brackets of (17) will contain a term of the form $(\alpha_1\beta_t)^2$ and consequently $E(\Sigma_2^2) > 0$, and for N fixed will increase as $(n-N)$.

Next we have
$$\Sigma_3 = \Sigma\Sigma A_{ij}\eta_i\zeta_j,$$
where either
$$i \leqslant n,\ j > n \quad \text{and} \quad j-i < N, \tag{18}$$
or
$$j \leqslant n,\ i > n \quad \text{and} \quad i-j < N, \tag{19}$$
and
$$\begin{aligned} A_{ij} &= \alpha_{p+1}\beta_1 + \ldots + \alpha_N\beta_{N-p} \quad \text{for} \quad i-j = p > 0 \\ &= \alpha_1\beta_{p+1} + \ldots + \alpha_{N-p}\beta_N \quad \text{for} \quad j-i = p > 0. \end{aligned}$$
Then
$$E(\Sigma_3) = 0, \quad E(\Sigma_3^2) = \sigma_1^2\sigma_2^2\Sigma\Sigma A_{ij}^2,$$
where the sum is taken over the values (18) and (19).

Finally
$$\Sigma_4 = \sum_{i=n+1}^{n+N}\sum_{j=n+1}^{n+N} A_{ij}\eta_i\zeta_j,$$
where
$$\begin{aligned} A_{ij} &= \alpha_{i-n}\beta_{i-p-n} + \ldots + \alpha_N\beta_{N-p} \quad \text{for} \quad i-j = p > 0 \\ &= \alpha_{j-p-n}\beta_{j-n} + \ldots + \alpha_{N-p}\beta_N \quad \text{for} \quad j-i = p > 0 \\ &= \alpha_p\beta_p \qquad + \ldots + \alpha_N\beta_N \quad \text{for} \quad i = j = n+p > n, \end{aligned}$$
and
$$E(\Sigma_4) = 0, \quad E(\Sigma_4^2) = \sigma_1^2\sigma_2^2 \sum_{i=n+1}^{n+N}\sum_{j=n+1}^{n+N} A_{ij}^2.$$
We readily see that
$$E(\Sigma_i\Sigma_j) = 0 \quad \text{for} \quad i \neq j$$
and therefore
$$E(S_n'^2) = E(\Sigma_1^2) + E(\Sigma_2^2) + E(\Sigma_3^2) + E(\Sigma_4^2).$$
Moreover, for constant N, $E(\Sigma_1^2)$, $E(\Sigma_3^2)$ and $E(\Sigma_4^2)$ are constant, and so for large n we have
$$n^{-1}E(S_n'^2) \to C_2 = \sigma_1^2\sigma_2^2[(\alpha_1\beta_N)^2 + (\alpha_1\beta_{N-1} + \alpha_2\beta_N)^2 + \ldots \\ + (\alpha_1\beta_1 + \ldots + \alpha_N\beta_N)^2 + \ldots + (\alpha_N\beta_1)^2] \neq 0. \tag{20}$$

Now suppose that N is fixed and consider the sum $\sum_1^n X_1(-t)\,Y_1(-t)$. We write
$$n = m(m+N) + p,$$
where $p < 2m + N + 1$ and n is large enough for m to be greater than N. This equation fixes m which increases roughly as $n^{\frac{1}{2}}$ when n increases. Write
$$\begin{aligned} S_n' &= \left(\sum_{t=1}^{N} + \sum_{N+1}^{N+m} + \ldots + \sum_{1+m(N+m-1)}^{m^2+mN} + \sum_{n-p+1}^{n}\right) X_1(-t)\,Y_1(-t) \\ &= V_1 + U_1 + V_2 + U_2 + \ldots + V_m + U_m + W. \end{aligned}$$

Then $V_1, \ldots, V_m$ and W are all independent and $E(V_1^2), \ldots, E(V_m^2)$ are independent of n, and in fact not greater than KN, where K is a constant independent of N. Also $E(W^2)$ is not greater than $K(2m + N + 1)$, where K may be taken as the same constant. $U_1, \ldots, U_m$ are also all independent and $E(U_m^2)$ is asymptotically equal to mC_2 when n (and therefore m) are large. Therefore, writing
$$A_m = U_1 + \ldots + U_m, \quad B_m = V_1 + \ldots + V_m + W,$$

we have
$$E(A_m) = 0, \quad E(A_m^2) = \sum_{i=1}^{m} E(U_i^2),$$
$$E(B_m) = 0, \quad E(B_m^2) = \sum_{i=1}^{m} E(V_i^2) + E(W^2),$$
and the latter increases as m, whilst the former increases as m^2 and so
$$E(B_m^2)\{E(A_m^2)\}^{-1} \to 0$$
as n increases.

By Lemma I it is therefore sufficient to show that the distribution of A_m tends to normality.

LEMMA II (Liapounoff's Central Limit Theorem, Bernstein, 1927). If
$$\Sigma_m = u_1^{(m)} + \ldots + u_m^{(m)}$$
is the sum of m independent quantities such that
$$E(u_r^{(m)}) = 0, \quad E(u_r^{(m)2}) = b_r^{(m)}, \quad E(u_r^{(m)4}) = c_r^{(m)},$$
and if, as m increases,
$$b_m^{-2} \sum_{r=1}^{m} c_r^{(m)} \to 0,$$
where
$$b_m = \sum_{r=1}^{m} b_r^{(m)} = E(\Sigma_m^2),$$
then
$$pr\{t_0(2b_m)^{\frac{1}{2}} \leqslant \Sigma_m < t_1(2b_m)^{\frac{1}{2}}\}$$
tends, uniformly in t_0 and t_1, to
$$\pi^{-\frac{1}{2}} \int_{t_0}^{t_1} e^{-t^2}\, dt.$$

To apply the lemma we put $U_r = u_r^{(m)}$. We already have $E(U_r) = 0$. Also
$$m^{-1}E(U_r^2) \to C_2 > 0 \quad \text{by (20)},$$
and so
$$m^{-2}b_m \to C_2.$$
Now consider
$$c_r^{(m)} = E(U_r^4),$$
where
$$U_r = \sum_{p=1}^{\infty} \sum_{q=1}^{\infty} A_{pq}\eta_{r(N+r-1)+p-1}\zeta_{r(N+r-1)+q-1},$$
and the A_{pq} are calculated with m in place of n. Since the η's all have the same probability distribution and similarly for the ζ's, we shall write η_p and ζ_q for $\eta_{r(N+r-1)+p-1}$ and $\zeta_{r(N+r-1)+q-1}$ for the sake of convenience. So we can write the above
$$U_r = \sum_{p=1}^{\infty} \sum_{q=1}^{\infty} A_{pq}\eta_p\zeta_q.$$

U_r^4 will be a polynomial of the fourth order in the η's and the ζ's and its expectation may be regarded as the sum of two distinct types of terms so that $E(U_r^4) = \Sigma E(w_1) + \Sigma E(w_2)$, where the terms w_1 are of the form $A_{pq}^4\eta_p^4\zeta_q^4$, and the terms w_2 are of the form $A_{pq}^2 A_{kl}^2 \eta_p^2\zeta_q^2\eta_k^2\zeta_l^2$ with $(p,q) \neq (k,l)$. All other terms arising in the product will clearly vanish when the expectation is taken.

Then, since the A_{pq} are bounded and the number of non-zero terms in w_1 and w_2 are not greater than $2N(m+N)$ and $4N^2(m+N)^2$ respectively, we have
$$E(U_r^4) < Km^2,$$
where K is a constant depending on N but independent of m and n. It follows that
$$b_m^{-2} \sum_{1} C_r^{(m)}$$

is of order m^{-1} and tends to zero as n and m increase. The conditions of the lemma are therefore satisfied and we conclude that

$$pr\{t_0(2E(A_m^2))^{\frac{1}{2}} \leqslant A_m < t_1(2E(A_m^2))^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to $$\pi^{-\frac{1}{2}} \int_{t_0}^{t_1} e^{-t^2}\, dt.$$

Applying Lemma I we have

$$pr\{t_0[2E(S_n'^2)]^{\frac{1}{2}} \leqslant S_n' < t_1[2E(S_n'^2)]^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to the same limit.

We now consider the relationship between S_n' and S_n. Write

$$\begin{aligned} S_n'' = S_n - S_n' &= \sum_1^n X(-t)\, Y(-t) - \sum_1^n X_1(-t) Y_1(-t) \\ &= \sum_{t=1}^n \left(\sum_{i=1}^\infty \alpha_i \eta_{t+i}\right)\left(\sum_{j=1}^\infty \beta_j \zeta_{t+j}\right) - \sum_{t=1}^n \left(\sum_{i=1}^N \alpha_i \eta_{t+i}\right)\left(\sum_{j=1}^N \beta_i \zeta_{t+j}\right) \\ &= \sum_{t=1}^n \left(\sum_{i=N+1}^\infty \alpha_i \eta_{t+i}\right)\left(\sum_{j=N+1}^\infty \beta_j \zeta_{t+j}\right) \\ &\quad + \sum_{t=1}^n \left(\sum_{i=N+1}^\infty \alpha_i \eta_{t+i}\right)\left(\sum_{j=1}^N \beta_j \zeta_{t+j}\right) \\ &\quad + \sum_{t=1}^n \left(\sum_{i=1}^N \alpha_i \eta_{t+i}\right)\left(\sum_{j=N+1}^\infty \beta_j \zeta_{t+j}\right) \\ &= W_1 + W_2 + W_3. \end{aligned} \tag{21}$$

We must now calculate the variance of these terms. Consider again (9). We have shown (11) that

$$\begin{aligned} 2\sum_1^\infty c_s d_s &\leqslant 2\sigma_1^2\sigma_2^2(|\beta_0| + |\beta_1| + \ldots)^2 \\ &\leqslant 2\sigma_1^2\sigma_2^2(|\alpha_0| + |\alpha_1| + \ldots)^2, \end{aligned}$$

and we now apply this to the three sums in equation (21). It follows that if N be chosen to satisfy the conditions (12) and (13) then

$$\overline{\lim}\, n^{-1}E(W_1^2) < K\sigma_1^2\sigma_2^2\epsilon^2, \quad \overline{\lim}\, n^{-1}E(W_2^2) < K\sigma_1^2\sigma_2^2\epsilon^2, \quad \overline{\lim}\, n^{-1}E(W_3^2) < K\sigma_1^2\sigma_2^2\epsilon^2,$$

where K is a constant independent of N.

It follows that $$S_n = S_n' + W_1 + W_2 + W_3,$$

where the variance of W_1, W_2 and W_3 can be made small compared with that of S_n by choosing N large. Then by first choosing N large and then n and using Tchebycheff's inequality, we see that the distribution of S_n tends to normality with variance $E(S_n^2)$ and this completes the proof.

In the general application of the above results some care is needed. We can suppose that our empirical values of X and Y are distributed about their sample means which we take to be zero and we must estimate the variance of S_n from formulae (9), or (approximately) from (10). But we must not insert in this formula the sample covariances for the c_s and the d_s because, as Bartlett (1946) has shown, the standard errors of the sample values of these covariances are of order $n^{-\frac{1}{2}}$ and we cannot therefore expect the series (10) to converge, let

alone give the correct value. To use the formula correctly we must first decide on the order and coefficients of the stochastic difference equation which we can suppose generated the series and, from these coefficients, calculate the value of (9).

In the case where the series are generated by a three-term difference equation, the calculations are simplified. Suppose the X and Y satisfy the equations

$$X(t+2)+aX(t+1)+bX(t)=\eta(t+2),$$
$$Y(t+2)+AY(t+1)+BY(t)=\zeta(t+2),$$

where
$$E(\eta(t))=E(\zeta(t))=0$$

and
$$E(\eta^2(t))=\sigma_1^2,\quad E(\zeta^2(t))=\sigma_2^2,$$

as before. For the series to be stationary, we must have $b<1$, $B<1$. We suppose that in addition to this the series are oscillatory and so $a^2<4b$, $A^2<4B$. The solutions will then be

$$X(t)=\sum_{s=0}^{\infty}2(4b-a^2)^{-\frac{1}{2}}p^s\sin\theta s\,\eta(t-s+1),$$

$$Y(t)=\sum_{s=0}^{\infty}2(4B-A^2)^{-\frac{1}{2}}P^s\sin\phi s\,\zeta(t-s+1),$$

where $p=b^{\frac{1}{2}}$, $P=B^{\frac{1}{2}}$, $\cos\theta=-a(2b^{\frac{1}{2}})^{-1}$, $\cos\phi=-A(2B^{\frac{1}{2}})^{-1}$. Also (Kendall, 1946, p. 408)

$$r_s=\frac{c_s}{c_0}=\frac{p^s\sin(s\theta+\psi)}{\sin\psi},\quad R_s=\frac{d_s}{d_0}=\frac{P^s\sin(s\phi+\Phi)}{\sin\Phi},$$

where
$$\tan\psi=\frac{1-p^2}{1+p^2}\tan\theta\quad\text{and}\quad\tan\Phi=\frac{1-P^2}{1+P^2}\tan\phi$$

and
$$c_0=\sigma_1^2\frac{1+b}{(1-b)\{(1+b)^2-a^2\}},\quad d_0=\sigma_2^2\frac{1+B}{(1-B)\{(1+B)^2-A^2\}}.$$

We then need to calculate

$$\begin{aligned}C&=c_0d_0+2\sum_{s=1}^{\infty}c_sd_s\\&=c_0d_0\left\{1+2\sum_{s=1}^{\infty}r_sR_s\right\}\\&=c_0d_0\left\{1+\sum_{s=1}^{\infty}\frac{p^sP^s\sin(s\theta+\psi)\sin(s\phi+\Phi)}{\sin\psi\sin\Phi}\right\}\\&=c_0d_0\left\{1+\frac{2pP}{\sin\psi\sin\Phi}\left[\frac{\cos(\psi-\Phi+\theta-\phi)-pP\cos(\psi-\Phi)}{1-2pP\cos(\theta-\phi)+p^2P^2}\right.\right.\\&\qquad\left.\left.-\frac{\cos(\psi+\Phi+\theta+\phi)-pP\cos(\psi+\Phi)}{1-2pP\cos(\theta+\phi)+p^2P^2}\right]\right\}.\end{aligned}\tag{22}$$

It is probably easiest to calculate C from this equation rather than attempt to simplify (22) still further. I hope to discuss the practical application of these formulae in another paper.

REFERENCES

WOLD, H. (1938). *A Study in the Analysis of Stationary Time Series.* Uppsala.
BERNSTEIN, S. (1927). *Math. Ann.* **97**, 1.
BARTLETT, M. S. (1946). *Supp. J. Roy. Statist. Soc.* **8**, 27.
KENDALL, M. G. (1946). *Advanced Theory of Statistics*, 2. London: Charles Griffin and Co.

RANK CORRELATION BETWEEN TWO VARIABLES, ONE OF WHICH IS RANKED, THE OTHER DICHOTOMOUS

By J. W. WHITFIELD, *Psychological Laboratory, University of Cambridge*

Rank correlation is one of the most useful statistical techniques available for the treatment of data arising in experimental and applied psychological research. Chambers (1946) has indicated the type of data most frequently occurring in these fields, and has pointed out the advantages of Kendall's τ over Spearman's ρ or any form of transformation to ordinal form.

Given the use of τ when tied rankings are present (Kendall, 1946) it seemed possible to extend the method to cover a very common problem in psychology, namely, determination of the relation between two variables, one of which is expressed as a ranking and the other as a dichotomy. In applied or field work the relation of a psychological 'measurement' and an external criterion nearly always appears in this form. The usual method of determining the relationship consists of reducing the ranking to a dichotomy and calculating χ^2 for the 2×2 table which results. That this may lead to inaccuracy can be seen from the following example:

Variable A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Variable B	+	+	+	+	+	+	+	−	−	−	+	+	+	−	−	−	−	−	−	−
Variable C	−	−	−	+	+	+	+	+	+	+	−	−	−	−	−	−	−	+	+	+

Here the data are supposed to be ranked according to variable A and dichotomized into + and − with respect to variables B and C.

Treating the relation between variables A and B as a 2×2 contingency table:

	Variable B	
	+	−
Variable A: Rankings 1–10	7	3
Rankings 11–20	3	7

Applying χ^2, P is found to be 0·074 without Yates's correction for continuity, or 0·180 if the correction is applied.

But χ^2 is exactly the same for the contingency table relating variables A and C, although it is obvious from the data that there is considerable difference in the two relationships, the evidence for which is sacrificed by reducing the ranking to a dichotomy.

If, alternatively, we consider the dichotomous variable as a ranking composed entirely of two sets of tied rankings, we may calculate the coefficients between A and B, A and C respectively which I shall denote by τ_{AB}, τ_{AC}. The corresponding values of S will be found to be, after the manner described by Kendall (1946):

$$S = +70 - 9 + 21 = +82,$$

$$S = -30 + 49 - 21 = -\ 2.$$

For the calculation of τ in the case of tied rankings we have a choice in the denominator by which S is to be divided to give τ. In the untied case this would be $\frac{1}{2}n(n-1)$, where n is

the number of ranks. In the tied case we may take the denominator S' as $\frac{1}{2}n(n-1)$ or as $[\frac{1}{2}n(n-1)\{\frac{1}{2}n(n-1)-\frac{1}{2}\Sigma t(t-1)\}]^{\frac{1}{2}}$, where t_1, t_2, etc., are the extent of the ties. The choice is determined by practical considerations (see Kendall, 1946), but is not material to a discussion of significance. For an untied ranking and a dichotomy with 'x' and 'y' members in each class, the second form reduces to $\{\frac{1}{2}xyn(n-1)\}^{\frac{1}{2}}$.

In the case of two untied rankings Kendall has shown that var $S = \frac{1}{18}n(n-1)(2n+5)$. In the case of one untied ranking and one with ties of extent t_1, t_2, etc., Sillitto (1947) has extended this result by proving that

$$\operatorname{var} S = \tfrac{1}{18}\{n(n-1)(2n+5)-\Sigma t(t-1)(2t+5)\}. \tag{1}$$

In the case of an untied ranking and a dichotomy, $t_1 = x$, $t_2 = y$, $(x+y) = n$, and we have then the simple form

$$\operatorname{var} S = \tfrac{1}{3}xy(n+1). \tag{2}$$

In the example above this gives

$$\operatorname{var} S = \frac{(10)(10)(21)}{3} = 700, \quad \sqrt{(\operatorname{var} S)} = 26{\cdot}46, \quad \frac{S_{AB}}{\sqrt{(\operatorname{var} S)}} = \frac{82-1}{26{\cdot}46} = 3{\cdot}06.$$

The probability of a deviation greater than this in absolute value is 0·0022. Further,

$$\frac{S_{AC}}{\sqrt{(\operatorname{var} S)}} = \frac{2-1}{26{\cdot}46} = 0{\cdot}0378,$$

and the corresponding probability is 0·970.

VARIANCE WHEN THERE ARE TIES IN THE RANKING

The variance of S given by equation (2) is true only in the case of a dichotomy and an untied ranking. For a tied ranking I surmised from some special cases that

$$\operatorname{var} S = \frac{xy}{3n(n-1)}\{(n^3-n)-\Sigma(t^3-t)\}. \tag{3}$$

In the note following this paper Mr Kendall provides proof of this result.

Example (from data collected by the Medical Research Council team in Germany 1946, as yet unpublished). Selected workers in a factory were interviewed and an assessment made of their adaptation to living conditions. They were assessed as 'Efficient' or 'Overactive'. Other data were available, including statements by the men of frequency of nocturia. For men aged 50–59 years the following was observed:

Assessment	Rank order of frequency of nocturia (least frequent nocturia given highest rank)
Efficient	$2\frac{1}{2}$, $2\frac{1}{2}$, $2\frac{1}{2}$, $2\frac{1}{2}$, $6\frac{1}{2}$, $6\frac{1}{2}$, 10, 10, 10, 10, 14, 14
Overactive	5, 10, 14, 16, 17

Five is the highest ranking in the overactive group. Four members of the efficient group have higher rankings, and eight lower rankings. The S score for that member is therefore $4-8$. Similarly, for all members we have

$$S = 4-8+6-2+10+12+12 = +34.$$

Using a denominator in the form

$$[xy\{\tfrac{1}{2}n(n-1)-\tfrac{1}{2}\Sigma t(t-1)\}]^{\frac{1}{2}},$$

τ is given by

$$\tau = +\frac{34}{\sqrt{[(12)(5)\{\frac{1}{2}(17)(16)-\frac{1}{2}(4)(3)-\frac{1}{2}(2)(1)-\frac{1}{2}(5)(4)-\frac{1}{2}(3)(2)\}]}} = +\frac{34}{\sqrt{6960}} = +0{\cdot}408.$$

From (3) we then have

$$\begin{aligned}\operatorname{var} S &= \frac{(12)(5)}{3(17)(16)}\{(17^3-17)-(4^3-4)-(2^3-2)-(5^3-5)-(3^3-3)\}\\ &= 344{\cdot}6.\end{aligned}$$

A small problem arises when we consider the correction for continuity to be applied in testing the significance of an observed value of S. In the case of a dichotomy and an untied ranking the interval between successive S values is 2. In the case of a dichotomy and a ranking composed entirely of ties of the same extent 't', the interval is $2t$. But in the example the ties are of varying extent, and the interval between successive S values is composed of a mixture of the intervals produced by the successive rank values. Thus, although these varying intervals are combined so that over most of the range the interval between successive values is unity, the distribution oscillates somewhat, and to use the value $\frac{1}{2}$ as the correction for continuity would sometimes be misleading. Further work is required to determine the correction which will provide a probability on the normal distribution equal to or slightly greater than the true probability in all cases. Until this is available I propose to use a crude correction, based on the average of the intervals mentioned above. In the example the successive rank values $2\frac{1}{2}$ and 5 give an interval of 5 in S score, rank values 5 and $6\frac{1}{2}$ give an interval of 3, and it is therefore possible to determine the average interval by calculating the intervals given by successive rank values. This calculation can be shortened. The total of the S score intervals is twice the number of members, less the extent of the ties involving the first and last members. If we divide this by the number of intervals between successive rank values we have the average S score interval. In the example this is $\frac{1}{6}(34-4-1)$. Using half of this as the correction for continuity we have

$$\frac{S}{\sqrt{(\operatorname{var} S)}} = \frac{34-2{\cdot}42}{18{\cdot}56} = 1{\cdot}702.$$

The pre-observational hypothesis, made on psychological grounds, was that excessive nocturia is a symptom of inefficient adaptation to living conditions, i.e. a positive correlation should be obtained. From these observations the probability of a positive correlation as great or greater than the observed value appearing by chance is 0·044. Direct calculation of the positive tail of the distribution of S gives a probability of 0·0368.

The alternative testing hypothesis based on the absolute value of S gives a probability twice as great, and the corresponding direct calculation using both positive and negative tails of the actual S distribution gives a probability of 0·0735.

By itself this evidence could only be debatable substantiation of the psychological hypothesis. In fact, additional data from two other factory groups, treated in the same way, gave a total S value of + 104, the square-root of the total variance being 35·00, providing a justification of the hypothesis.

THE CASE OF THE 2×2 TABLE

If one dichotomous variable can be considered as a ranking with two sets of tied ranks it is logical to consider the case when both variables are in this form. If we have a 2×2 table in the form

(AB)	(Ab)	(A)
(aB)	(ab)	(a)
(B)	(b)	N

any member of (AB) taken with any member of (ab) has the same order in either ranking and hence contributes $+1$ to S, and any member of (Ab) with any member of (aB) contributes -1. The others contribute nothing. Hence

$$S = (AB)(ab) - (Ab)(aB).$$

From equation (3)

$$\begin{aligned} \operatorname{var} S &= \frac{(A)(a)}{3N(N-1)}[(N^3 - N) - \{(B)^3 - (B)\} - \{(b)^3 - (b)\}] \\ &= \frac{(A)(a)(B)(b)}{N-1}. \end{aligned} \tag{4}$$

Again, for testing the significance of an observed value of S it is necessary to correct for continuity by subtracting half the interval between successive S values. In the case of the 2×2 table the interval is N, for if we increase (AB) by unity S becomes

$$\{(AB)+1\}\{(ab)+1\} - \{(Ab)-1\}\{(aB)-1\} = (AB)(ab) - (Ab)(aB) + N.$$

Hence, for the normal deviate, we have

$$\frac{S - \frac{1}{2}N}{\sqrt{\left\{\frac{(A)(a)(B)(b)}{N-1}\right\}}}. \tag{5}$$

It will be noted that τ (taking the ties into account in calculating the denominator S') is

$$\frac{(AB)(ab) - (Ab)(aB)}{\sqrt{[(A)(a)(B)(b)]}},$$

which is the product-moment correlation for a 2×2 table when the variables are conventionally regarded as possessing the discrete values 0, 1.

Testing by use of the normal deviate seems to be moderately accurate, and would appear to be useful in those cases where χ^2 is suspect because of small expectations in the cells of the 2×2 table. It is less laborious to calculate than the hypergeometric treatment, and is an alternative form of the approximation to hypergeometric treatment given by Pearson (1947), who also discusses the order of accuracy of the approximation.

Using the data given earlier as an example, but assuming that it had been possible only to grade nocturia into 'Normal' or 'Excessive', we have the following table:

Assessment	Nocturia	
	Normal	Excessive
Efficient	10	2
Overactive	2	3

$$S = 30-4 = 26, \quad \text{var}\, S = \frac{(12)(5)(12)(5)}{16} = 225.$$

This gives, after correction for continuity,

$$\frac{S-\frac{1}{2}N}{\sqrt{(\text{var}\, S)}} = \frac{26-8\cdot5}{15} = 1\cdot1667.$$

This gives the probability of S being attained or exceeded in the direction of the hypothesis (i.e. positive values only) as 0·1217. χ^2 without the continuity correction gives $P = 0\cdot0369$,* and with the correction, $P = 0\cdot1143$. The hypergeometric treatment, summing the probabilities of obtaining 3, 4 or 5 in the Overactive-Excessive category, gives $P = 0\cdot1166$.

If the more customary test of absolute value is applied, χ^2 with Yates's correction gives $P = 0\cdot2286$, S and the normal deviate gives $P = 0\cdot2434$, i.e. both values of P are doubled. The hypergeometric treatment, adding the probability of obtaining 0 in the Overactive-Excessive category gives $P = 0\cdot2445$.

It will be seen that in conditions such as these, S and the normal deviate give a reasonable approximation to the exact treatment.

* This is making the common assumption that $\{(AB)(ab)-(Ab)(aB)\}^2 N/\{(A)(a)(B)(b)\}$ is distributed as χ^2 with 1 degree of freedom, or that its square root is a normal deviate with sign depending on the sign of $(AB)(ab)-(Ab)(aB)$.

REFERENCES

CHAMBERS, E. G. (1946). Statistical techniques in applied psychology. *Biometrika*, **33**, 269.

KENDALL, M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**, 81.

KENDALL, M. G. (1946). The treatment of ties in ranking problems. *Biometrika*, **33**, 239.

PEARSON, E. S. (1947). The choice of statistical tests illustrated on the interpretation of data classed in a 2 × 2 table. *Biometrika*, **34**, 139.

SILLITTO, G. P. (1947). The distribution of Kendall's coefficient of rank correlation in rankings containing ties. *Biometrika*, **34**, 36.

THE VARIANCE OF τ WHEN BOTH RANKINGS CONTAIN TIES

BY M. G. KENDALL

1. The variance of τ in the population of sample permutations was given in my paper of 1938 for the case where no tied ranks exist. Mr Sillitto (1947) has given the formula where one ranking contains ties but the other does not. In the foregoing paper Mr Whitfield has correctly surmised the variance when one ranking contains ties, and the other is a dichotomy. In this note I derive the general formula for the variance when both rankings contain ties. The results of Messrs Sillitto and Whitfield then follow as special cases.

2. I shall follow the method of Daniels (1944). If a_{ij} represents the contribution of the ith and jth members of a ranking to τ we have

$$\begin{aligned} a_{ij} &= +1 \quad (i<j) \\ &= 0 \quad (i=j) \\ &= -1 \quad (i>j) \end{aligned} \tag{1}$$

We write
$$c_{ij} = a_{ij} b_{ij}, \tag{2}$$
where a and b refer to different rankings, and
$$c = \sum_{i,j=1}^{n} c_{ij}. \tag{3}$$
The quantity c is simply related to S by the relation
$$c = 2S, \tag{4}$$
and for the testing of τ it is sufficient to test c or S which are merely constant multiples of τ. I work with the quantity c.

3. We have, from Daniel's results,

$$\sum_{l=1}^{n} a_{il} = n+1-2i, \tag{5}$$

$$\sum_{i,l=1}^{n} a_{il}^2 = n(n-1), \tag{6}$$

$$\sum_{i,l,t=1}^{n} a_{il}\, a_{it} = \tfrac{1}{3}n(n^2-1), \tag{7}$$

$$E(c) = 0, \tag{8}$$

$$E(c^2) = \frac{4}{n(n-1)(n-2)}\{\Sigma a_{il} a_{it} - \Sigma a_{il}^2\}\{\Sigma b_{il} b_{it} - \Sigma b_{il}^2\} + \frac{2}{n(n-1)} \Sigma a_{il}^2 \Sigma b_{il}^2. \tag{9}$$

If we substitute from (6) and (7) in (9) we find
$$E(c^2) = \tfrac{2}{9}n(n-1)(2n+5), \tag{10}$$
or, equivalently,
$$E(S^2) = \tfrac{1}{18}n(n-1)(2n+5), \tag{11}$$
from which the variance of τ in the case of untied rankings follows at once.

4. Now suppose that sets of t_1, t_2, ... consecutive members in one ranking are tied. In place of (6) we then have
$$\sum_{i,l=1}^{n} a_{il}^2 = n(n-1) - \Sigma t(t-1), \tag{12}$$

the summation on the right taking place over the various values of t. This result follows simply from the consideration that for a pair of tied ranks $a_{ij} = 0$, and consequently the sum of squares of contributions from a tied set is of the same form as for the ranking as a whole.

In place of (7) we have

$$\sum_{i,l,t=1}^{n} a_{il} a_{it} = \tfrac{1}{3}n(n^2-1) - \tfrac{1}{3}\Sigma\, t(t^2-1). \tag{13}$$

This is not quite so obvious. Consider a set of tied ranks. The contribution to the sum on the left of (13) will be unchanged if the suffixes l, t fall outside this set. If they both fall inside, no contribution arises and therefore we have to subtract the term $\frac{1}{3}t(t^2-1)$. The remaining possibility is that one falls inside and one outside. In such a case the contribution remains unchanged in total for it is zero in the original untied case, each possible pair occurring once to give $+1$ and one to give -1. Formula (13) follows.

5. By substitution in (9) we then have, for two rankings with ties typified respectively by t and u,

$$\begin{aligned} E(c^2) = {} & \frac{4}{n(n-1)(n-2)}\{\tfrac{1}{3}n(n-1)(n-2) - \tfrac{1}{3}\Sigma t(t-1)(t-2)\} \\ & \times \{\tfrac{1}{3}n(n-1)(n-2) - \tfrac{1}{3}\Sigma u(u-1)(u-2)\} \\ & + \frac{2}{n(n-1)}\{n(n-1) - \Sigma t(t-1)\}\{n(n-1) - \Sigma u(u-1)\}. \end{aligned} \tag{14}$$

This is the general formula required. We can express it in the alternative form

$$\begin{aligned} E(c^2) = {} & \tfrac{2}{9}n(n-1)(2n+5) - \tfrac{2}{9}\Sigma t(t-1)(2t+5) - \tfrac{2}{9}\Sigma u(u-1)(2u+5) \\ & + \frac{4}{9n(n-1)(n-2)}\{\Sigma t(t-1)(t-2)\}\{\Sigma u(u-1)(u-2)\} \\ & + \frac{2}{n(n-1)}\{\Sigma t(t-1)\}\{\Sigma u(u-1)\}. \end{aligned} \tag{15}$$

6. (i) If one ranking is untied, say all the u's are zero, we have Mr Sillitto's result

$$E(c^2) = \tfrac{2}{9}n(n-1)(2n+5) - \tfrac{2}{9}\Sigma t(t-1)(2t+5). \tag{16}$$

(ii) If one ranking is untied and the other is a dichotomy into x and $n-x=y$ members, (16) reduces to

$$E(c^2) = \tfrac{4}{3}xy(n+1), \tag{17}$$

agreeing with Mr Whitfield's equation (2).

(iii) If one ranking contains ties and the other is a dichotomy we find on substitution in (14)

$$E(c^2) = \frac{4xy}{3n(n-1)}\{n^3 - n - \Sigma(t^3-t)\}, \tag{18}$$

agreeing with Mr Whitfield's equation (3).

(iv) Finally, if both variates are dichotomized into x, y and p, q we find

$$E(c^2) = \frac{4xypq}{n-1}, \tag{19}$$

agreeing with Mr Whitfield's equation (4).

REFERENCES

See the references to Mr Whitfield's paper together with:

DANIELS, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika*, **33**, 129.

A χ^2 'SMOOTH' TEST FOR GOODNESS OF FIT

By F. N. DAVID

1. The χ^2 test occupies a central position in statistical theory, and it is difficult to imagine another test which would have the same generality of application. We shall be concerned here with one aspect only, that is, the uses of χ^2 in the tests for agreement between hypothesis and observation which are usually loosely classed together under the name of tests for goodness of fit. The principal advantages of χ^2 for such tests would seem to be (i) that it is applicable to grouped observations, (ii) that the parameters of the hypothesis tested may be calculated from the observational data and the fact allowed for in the degrees of freedom with which the criterion is assumed to be distributed and (iii) that it is easy to calculate, for the number of computations involved is just the number of groups into which the observational data are divided. It has, however, two defects which have long been known and which are easily recognized from the form of the criterion itself. Broadly speaking we may define χ^2 as follows:

$$\chi^2 = \text{sum for all groups of } \frac{(\text{Observed value} - \text{Expected value})^2}{\text{Expected value}}.$$

It will be seen (iv) that in taking the square of (Observed value − Expected value) the knowledge regarding the sign of the deviation is lost. Further (v) there is no means of preserving the order of the signs of the deviations, and no distinction can therefore be made between a departure from the hypothesis tested in which the deviations were first all positive and then all negative and a departure in which the sizes and signs of the deviations were random between themselves.

2. The ideal test for goodness of fit should certainly take into account (i), (ii), (iv) and (v) and probably (iii) also, but it is unlikely that this ideal will be reached. It would seem at the present time that the most which may be hoped for are tests which will supplement the χ^2 test in that they will be more sensitive to given alternatives of the hypothesis under test. Neyman (1937) put forward such a supplementary test in which he developed the ψ^2 criterion. This criterion was designed to be sensitive to alternate hypotheses of a type he designated as smooth; that is to say, if the hypothesis under test is a continuous curve, such as, for example, the normal curve, admissible hypotheses alternate to it might be other normal curves with a different mean and a different standard deviation. Neyman's criterion certainly took into account (iv) and (v), but it did not fulfil (i), in that it was only applicable to ungrouped observations,* or (ii), in that the parameters of the hypothesis tested had to be known *a priori*. Whether it fulfilled (iii) is a matter for personal opinion.

3. It would appear possible that tests which would take into account the sign only of the deviations of observation from hypothesis, and the order of these signs, may be devised from simple combinatorial principles. Suppose that there are N observations which are divided into k groups. Let n_i ($i = 1, 2, \ldots, k$) be the actual number of observations falling

* Prof. Neyman tells me that his criterion has been adapted for the case of grouped observations but that he has not yet published this extension.

into the ith group, and let m_i $(i = 1, 2, \ldots, k)$ be the expected number. It is possible theoretically for χ^2 to be calculated for the case where

$$\sum_{i=1}^{k} n_i = N \neq \sum_{i=1}^{k} m_i,$$

but such cases must be rare in statistical practice. We shall overlook this case and will consider the case where the totals of observed and expected are made equal to one another with the resultant loss of one degree of freedom in the calculation of χ^2. If the totals agree then

$$\sum_{i=1}^{k} n_i - \sum_{i=1}^{k} m_i = \sum_{i=1}^{k} \delta_i = 0,$$

where $\delta_i = n_i - m_i$. In order that the sum of these δ's should be zero, at least one of them must be negative in sign, but which one of these δ's it will be would seem to be a matter of chance. It is on this fact that we shall base the first test criterion.

4. Suppose that we have a sequence of signs of which r_1 are positive and r_2 negative, where $r_1 + r_2 = r$, and $r_1 > 0$ and $r_2 > 0$. These signs are postulated to occur in a random order. Given such a sequence it is easy to record the number of sets of positive and negative signs. For example, if the sequence is

$$+ + - - + + - - + - + + + + -,$$

then $r = 15$, $r_1 = 9$, $r_2 = 6$, and there are four sets of positive signs and four sets of negative signs. In general there can be (α) t positive, t negative, or (β) t positive, $t+1$ negative, or (γ) $t+1$ positive and t negative sets of signs. If $T = 2t$ or $2t+1$ as required, we may ask what is the probability that given r_1 and r_2 such a number T of sets (alternately positive and negative) would have arisen through chance. This probability follows at once from Whitworth, *Choice and Chance*, Proposition XXV, viz.: 'The number of ways in which n indifferent things can be distributed in t different parcels (blank lots being inadmissible) is

$$(n-1)!/(t-1)!\,(n-t)!\text{'.*}$$

5. The total number of ways in which r_1 and r_2 elements can be arranged is

$$\frac{r!}{r_1!\,r_2!}.$$

We now require to enumerate the number of ways in which r_1 can be arranged to form t sets and r_2 to form t sets. To arrange r_1 in t sets is equivalent (vide Whitworth) to making $t-1$ breaks in a sequence of r_1 observations, and this may be done in

$$(r_1-1)!/(t-1)!\,(r_1-t)! \text{ ways,}$$

and similarly for r_2. It is not specified whether $+$ or $-$ should start the sequence, and hence the total number of ways in which a sequence $r_1 + r_2$ may be arranged in t sets each is

$$\frac{2(r_1-1)!\,(r_2-1)!}{(t-1)!\,(t-1)!\,(r_1-t)!\,(r_2-t)!}.$$

* Since I first thought of this method of attack I have found that the distribution of groups as given by me in § 5 has already been given by W. L. Stevens, *Ann. Eugen., Lond.*, **9**, 10, and by A. Wald and J. Wolfowitz, *Ann. Math. Statist.* **11**, 147. The probability function has been tabled by F. S. Smed and C. Eisenhart, *Ann. Math. Statist.* **14**, 66, but it is not in a form that I found suitable for my purposes. The probability function has been known for many years; what is interesting is the different uses to which it has been put.

The probability of $2t$ sets will be

$$P\{2t \mid r_1, r_2\} = \frac{2r_1!(r_1-1)!\,r_2!(r_2-1)!}{r!(t-1)!\,(t-1)!\,(r_1-t)!\,(r_2-t)!}, \tag{1}$$

and the probability of obtaining $(2t+1)$ sets will be

$$P\{2t+1 \mid r_1, r_2\} = P\{t \mid r_1, t+1 \mid r_2\} + P\{t+1 \mid r_1, t \mid r_2\} = P\{2t \mid r_1, r_2\}\left(\frac{r-2t}{2t}\right). \tag{2}$$

Hence given r_1, r_2 and T from a random sequence of positive and negative signs the probability of such a number of sets having arisen through chance may be calculated.

6. It is desired to use the probability of a given arrangement of signs in order to test a given hypothesis represented by a smooth probability law, bearing in mind that, if the given hypothesis is not true, then any alternative law is likely to be of a smooth type. Although no exact definition of a smooth alternative distribution has been made, it may be stated here that *smooth*, in the sense used by Neyman, will imply that the number of sets of signs will be small. For example, if the hypothesis tested is that observations follow a given normal curve, whereas in fact they have been drawn from a normal distribution identical with the first but with a smaller mean, then the differences between observation and expectation on the basis of the hypothesis tested may be expected to give a preponderance of positive signs below the sample mean and of negative signs above it; that is to say, if the difference in means is sufficient to offset the sampling fluctuations we should find a single set of positive signs followed by a single set of negative signs. If the true population is a normal curve with the same mean but with a larger standard deviation than that specified by the hypothesis tested, then there will be a tendency towards a set of positive signs, a set of negative signs, followed by a set of positive signs, although sampling fluctuations may not leave such a clear-cut answer. The more complex the alternative hypothesis the less chance there will be of detecting it.

7. With this objective in view it is proposed to take T, the number of sets of signs, as the test criterion, rejecting the hypothesis tested whenever, for a given r_1 and r_2, T is exceptionally small. This we do on the grounds that the existence of very few sets of signs suggests that the differences between observed and expected frequencies are not due to chance sampling fluctuations but to some systematic departure of the true probability law (assumed *smooth*) from hypothesis. In following this procedure we should reject the hypothesis if

$$P\{T \leqslant T_0\} = \sum_{T=2}^{T_0} P\{T \mid r_1 r_2\} \leqslant \epsilon,$$

where T_0 is the observed value of T and ϵ the significance level selected as appropriate. Exact probabilities are given in Table 1, and the application of the test is immediate.* There seems to be no reason why the test should not be applicable to both grouped and ungrouped observations, although the formulation of the hypothesis tested may be somewhat different in the two cases. Consider a sample which has been supposedly randomly

* An assumption implicit in the test would appear to be that for each χ^2 cell there is an equal chance of obtaining a positive or a negative deviation, that is, that there are sufficient numbers in each cell for the binomial to be closely approximated to by a normal curve. An extensive series of random sampling experiments has shown, however, that the divergence between theory and practice is not significant even when the probability of obtaining a positive is four times that of obtaining a negative. Hence while strictly the expectation in each cell of χ^2 should be 10 or over, it would seem that for practical purposes that the T test may be applied in all cases where the application of the χ^2 test is permissible.

Table 1. *Probability of obtaining a given number of sets, T.* $[T = 2t$ or $2t+1]$

The function tabled is $\dfrac{2(r_1-1)!\,(r_2-1)!}{(t-1)!\,(t-1)!\,(r_1-t)!\,(r_2-t)!}$ or $\dfrac{(r_1-1)!\,(r_2-1)!\,(r_1+r_2-2t)}{t!(t-1)!\,(r_1-t)!\,(r_2-t)!}$, according as T is even or odd.

$P\{T\}$ is obtained by dividing this function by the binomial term $\dfrac{r!}{r_1!\,r_2!}$.

r	r_1	r_2	$\dfrac{r!}{r_1!\,r_2!}$	$T=2$	3	4	5	6	7	8	9	10	11	12	13	14
2	1	1	2	2												
3	2	1	3	2	1											
4	3	1	4	2	2											
	2	2	6	2	2	2										
5	4	1	5	2	3											
	3	2	10	2	3	4	1									
6	5	1	6	2	4											
	4	2	15	2	4	6	3									
	3	3	20	2	4	8	4	2								
7	6	1	7	2	5											
	5	2	21	2	5	8	6									
	4	3	35	2	5	12	9	6	1							
8	7	1	8	2	6											
	6	2	28	2	6	10	10									
	5	3	56	2	6	16	16	12	4							
	4	4	70	2	6	18	18	18	6	2						
9	8	1	9	2	7											
	7	2	36	2	7	12	15									
	6	3	84	2	7	20	25	20	10							
	5	4	126	2	7	24	30	36	18	8	1					
10	9	1	10	2	8											
	8	2	45	2	8	14	21									
	7	3	120	2	8	24	36	30	20							
	6	4	210	2	8	30	45	60	40	20	5					
	5	5	252	2	8	32	48	72	48	32	8	2				
11	10	1	11	2	9											
	9	2	55	2	9	16	28									
	8	3	165	2	9	28	49	42	35							
	7	4	330	2	9	36	63	90	75	40	15					
	6	5	462	2	9	40	70	120	100	80	30	10	1			
12	11	1	12	2	10											
	10	2	66	2	10	18	36									
	9	3	220	2	10	32	64	56	56							
	8	4	495	2	10	42	84	126	126	70	35					
	7	5	792	2	10	48	96	180	180	160	80	30	6			
	6	6	924	2	10	50	100	200	200	200	100	50	10	2		
13	12	1	13	2	11											
	11	2	78	2	11	20	45									
	10	3	286	2	11	36	81	72	84							
	9	4	715	2	11	48	108	168	196	112	70					
	8	5	1287	2	11	56	126	252	294	280	175	70	21			
	7	6	1716	2	11	60	135	300	350	400	250	150	45	12	1	
14	13	1	14	2	12											
	12	2	91	2	12	22	55									
	11	3	364	2	12	40	100	90	120							
	10	4	1001	2	12	54	135	216	288	168	126					
	9	5	2002	2	12	64	160	336	448	448	336	140	56			
	8	6	3003	2	12	70	175	420	560	700	525	350	140	42	7	
	7	7	3432	2	12	72	180	450	600	800	600	450	180	72	12	2

drawn from some population. Let the elements of the sample in order of drawing be $x_1, x_2, \ldots, x_n$. We may use the T criterion to test the hypothesis of randomness, in the following way. If u_1 is the smallest value observed in the sample and u_n the largest, then if we exclude the trivial case when all the x's are equal it is easy to show that $u_1 < \bar{x} < u_n$, where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. If we now consider the deviations

$$x_i - \bar{x} = \delta x_i \quad \text{for} \quad i = 1, 2, \ldots, n,$$

there will be a series $\delta x_1, \delta x_2, \ldots, \delta x_n$, some of which quantities will be positive and some negative. The application of the T test is immediate, the admissible alternate hypotheses being that if the drawing of the sample is not at random then bias of the smooth kind is present.

8. As an illustration consider the following two cases:

Case I.	Expected frequency	10	25	35	75	155	155	75	35	25	10
	Observation	12	29	45	81	160	145	69	31	20	8
	Deviation	+	+	+	+	+	−	−	−	−	−
Case II.	Expected frequency	10	25	35	75	155	155	75	35	25	10
	Observation	12	23	45	66	161	160	69	36	20	8
	Deviation	+	−	+	−	+	+	−	+	−	−

In the first case $\chi^2 = 6{\cdot}94$ and in the second $\chi^2 = 6{\cdot}80$; in neither case would the hypothesis be rejected as inadequate by using the χ^2 criterion. The T criterion does, however, bring out the essential difference:

Case I. $\quad r_1 = 5, \quad r_2 = 5, \quad T_0 = 2 \quad$ and $\quad P\{T \leqslant T_0\} = \frac{2}{252}$.

Case II. $\quad r_1 = 5, \quad r_2 = 5, \quad T_0 = 8 \quad$ and $\quad P\{T \leqslant T_0\} = \frac{242}{252}$.

Using the T test we should be inclined to reject the first hypothesis in favour of a smooth alternative, while for the second case we should be inclined to agree with the conclusion drawn from the χ^2 test that the observational material is adequately described.

9. Sampling material is available whereby the theoretical distribution of T may be tested in practice. Neyman & Pearson (1928) took 208 samples, each of size 200, from a population of eight groups described by the cubic curve

$$y = 25 + \tfrac{45}{8}x - \tfrac{15}{128}x^3.$$

The expectation in each cell for a sample of this size was calculated and the χ^2 criterion found for each of the 208 samples. The writer was given access to these calculations and was able to find the sampling distribution of T from the material. The results of this sampling experiment and the theoretical distribution of T from relations (1) and (2) are given in Table 2.

The agreement between theory and practice would seem to be reasonably good, and in the cases (4, 4) and (5, 3) the values of χ^2, calculated to test the discrepancy between theory and practice, were not greater than might be attributable to sampling fluctuations. It was not thought worth while to calculate χ^2 for (6, 2) and (7, 1). A second sampling experiment in which samples of size 360 were drawn from a normal population of fifteen groups lent further support to the reasonableness of the theoretical distribution.

10. The T criterion will be a useful supplementary criterion to the χ^2, but because it takes account solely of the sign of a distribution and not of its magnitude it will probably only be useful when used in conjunction with χ^2. A test of significance which could combine both the probability levels of T and χ^2 would undoubtedly be more useful, and we may

therefore consider how this might be done. Unless the exact degree of dependence which exists between two variables is known it is usually only possible to obtain their joint distribution if they are independent. It would appear reasonable, both on theoretical grounds and from sampling experiments, to assume that T and χ^2 are independent, or, if the assumptions underlying both tests are not exactly fulfilled, to assume that the degree of dependence between them is at most small.

Table 2. *Comparison of theoretical distribution of T with that derived from a sampling experiment*

(4 positive, 4 negative)

T = number of sets	2	3	4	5	6	7	8	Total
Sampling	3	5	20	25	28	5	7	93
Theory	2·7	8·0	23·9	23·9	23·9	8·0	2·7	93

(5 positive, 3 negative) or (3 positive, 5 negative)

T = number of sets	2	3	4	5	6	7	Total
Sampling	2	8	32	30	20	10	102
Theory	3·6	10·9	29·1	29·1	21·9	7·3	102

(6 positive, 2 negative) or (2 positive, 6 negative)

T = number of sets	2	3	4	5	Total
Sampling	1	3	—	5	9
Theory	0·6	1·9	3·2	3·2	9

(7 positive, 1 negative) or (1 positive, 7 negative)

T = number of sets	2	3	Total
Sampling	—	2	2
Theory	0·5	1·5	2

11. We shall begin by demonstrating that as far as mathematics are concerned the T and χ^2 criteria are completely independent.* For simplicity of argument let us consider the case of three groups only. The sample may then be represented by a point (n_1, n_2, n_3) in three-dimensioned space, with axes of reference On_1, On_2, On_3, and the expected population values by a point (m_1, m_2, m_3) in the same space. Since

$$n_1+n_2+n_3 = m_1+m_2+m_3 = N,$$

* This method of approach was suggested to me by Andrew Gleason of Harvard University at a seminar given at the Statistical Laboratory, University of California, at Berkeley.

these points are constrained to lie in a plane. Fig. 1 shows this plane for the particular case $N = 16$; $m_1 = 4$, $m_2 = 8$, $m_3 = 4$. Since no frequency can be negative, possible sample points must be within an equilateral triangle lying in this plane, the chance of occurrence associated with a point being the multinomial term

$$\frac{N!}{n_1!\,n_2!\,n_3!}\left(\frac{m_1}{N}\right)^{n_1}\left(\frac{m_2}{N}\right)^{n_2}\left(\frac{m_3}{N}\right)^{n_3}.$$

When using the χ^2 test the mathematical approximation consists in substituting for this term an expression proportional to $e^{-\frac{1}{2}\chi^2}$, in regarding this last as a continuous function, and

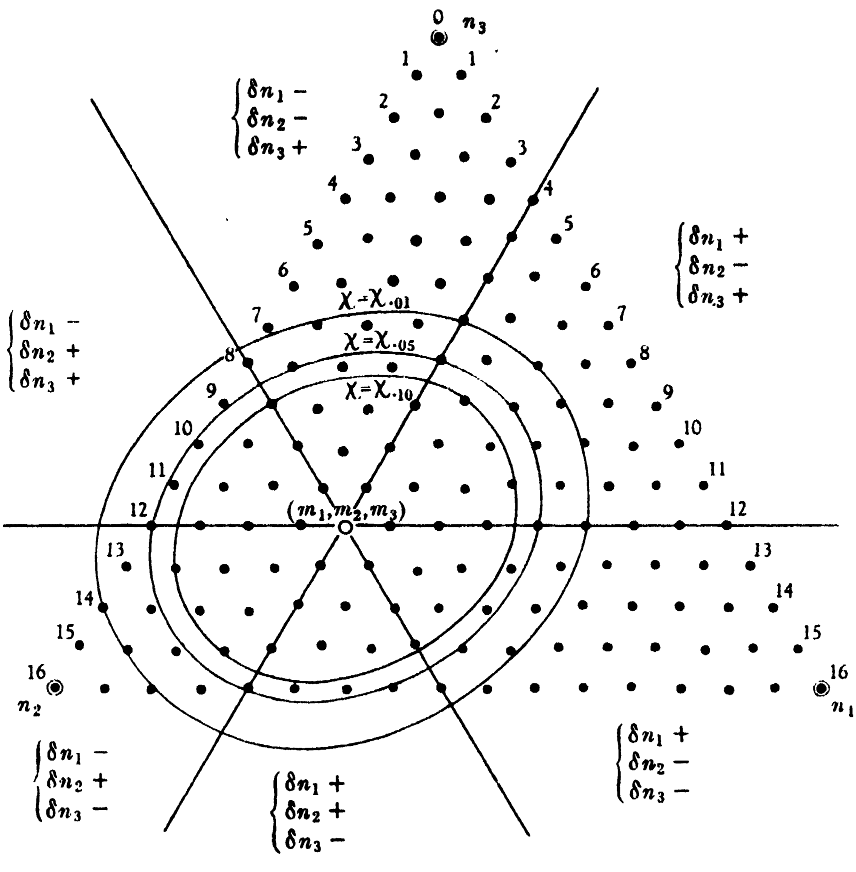


Fig. 1. Graphical illustration of the χ^2 contours and the change in signs of the δn's. n_1, n_2 and n_3 denote the points of intersection of the On_1, On_2, On_3 axes with the plane $n_1+n_2+n_3 = N$. According to the approximation, the chance equals α of obtaining a sample point lying outside the elliptic contour on which $\chi = \chi_\alpha$.

in taking as a measure of goodness of fit the integral of this expression outside the ellipse which passes through the sample point and on which χ^2 is constant. For the case of three groups this integral itself assumes the simple form $e^{-\frac{1}{2}\chi^2}$. Three such elliptic contours are shown in the diagram.

Planes through (m_1, m_2, m_3) parallel to the co-ordinate planes n_1On_2, n_2On_3, n_3On_1, will intersect the sample plane

$$n_1+n_2+n_3 = N$$

in three straight lines. As shown in the diagram, these lines divide the sample plane into six sectors, and for all sample points within a sector the signs of the differences $\delta n_i = n_i - m_i$ will remain unchanged. Any test based solely on runs of signs will consist in taking one or more of these sectors as critical regions and rejecting the hypothesis tested when the sample point falls therein. It is clear that if we use the mathematical approximation, the distribution of χ^2 is the same within each sector; similarly, that the chance of a sample showing a given combination of signs is the same on each ellipse along which χ^2 is constant. Thus under the assumptions made regarding the distribution of χ^2, the T and χ^2 criteria are completely independent.

In this case of three groups T can only assume values of two or three and the former value would not be judged significant, but the argument will follow exactly similar lines in the case of many groups. The number of sectors will be in general $2(2r_2+1)$ if $r_1 > r_2$ and $2(2r_2)$ if $r_1 = r_2$, and they will be bounded by primes passing through the population point.

12. While the distributions of T and χ^2 are independent for this mathematical model they are unlikely to be exactly so when we go back to the true multinomial density distribution, because the sample space is neither continuous nor infinite. The model, in fact, becomes inaccurate if m_1, m_2 or m_3 are very small. For example, it is seen in Fig. 1 that while the 1 % ellipse ($\chi = \chi_{0 \cdot 01}$) lies completely within the triangular space for the sectors with signs $+ - +$ and $- - +$, it lies completely without the space for the sector $+ + -$ and partly without for the other sectors. It has been thought worth while therefore to test whether the two criteria are independent in practice, and to this end the same material previously described has been utilized. Tables 3 and 4 give the distribution of mean χ^2 for different values of $P\{T\}$ and the distribution of mean $P\{T\}$ for grouped values of χ^2. There is little evidence in these figures to show that $P\{T\}$ and χ^2 (and therefore $P\{\chi^2\}$) are related. The figures therefore lend support to the geometrical argument and indicate that the approximations involved in χ^2, both from the small sample and the fact that the sample space is not infinite, do not invalidate the mathematical result.

13. In order to combine the χ^2 and T tests of significance it will be necessary to develop a theory for the combination of two tests of significance when one criterion is a continuous and the other a discontinuous variable. R. A. Fisher has set out the test for the combination of tests of significance from a number of independent continuous variables. The keystone of the test is the recognition of the fact that if Z is a continuous variable, then z, where

$$z = \int_{-\infty}^{Z} p(Z)\,dZ,$$

is also a continuous variable equally likely to have any value between 0 and 1; we shall describe z as being distributed rectangularly. Twice the logarithm of the product of two such z's, say z_1 and z_2, where z_1 and z_2 follow from two independent tests of significance can be shown to be distributed as χ^2 with four degrees of freedom. Consider a discontinuous variable X which may take values $X_1, X_2, \ldots, X_s$ and which has an elementary probability law $P\{X = X_j\} = p_j$, where

$$0 < p_j < 1 \quad \text{for} \quad j = 1, 2, \ldots, s$$

and

$$\sum_{j=1}^{s} p_j = 1.$$

If a new variable, x, is defined as taking values $x_1, x_2, \ldots, x_s$, where

$$x_k = \sum_{j=1}^{k} p_j,$$

Table 3. *Mean χ^2 for different values of $P\{T\}$*

r_1, r_2 or r_2, r_1	—	4, 4	5, 3	4, 4	5, 3	6, 2	4, 4	5, 3	4, 4	6, 2
No. of obs. on which mean is based	26	5	20	28	30	—	25	32	20	3
$P\{T\}$	1·00	0·97	0·93	0·86	0·71	0·64	0·63	0·43	0·37	0·29
Mean χ^2	7·19	6·04	5·87	6·43	5·98	—	7·41	7·32	6·82	4·86

r_1, r_2 or r_2, r_1	7,1	5, 3	4, 4	6, 2	5, 3 and 4, 4
No. of obs. on which mean is based	—	8	5	1	5
$P\{T\}$	0·25	0·14	0·11	0·07	0·04 and 0·03
Mean χ^2	—	6·10	5·53	6·95	8·06

Table 4. *Mean $P\{T\}$ for grouped χ^2*

χ^2	0·0–1·0	1·0–2·0	2·0–3·0	3·0–4·0	4·0–5·0	5·0–6·0	6·0–7·0	7·0–8·0	8·0–9·0	9·0–10·0
No. of obs. on which mean is based	1	7	14	27	32	29	21	14	11	15
Mean $P\{T\}$	0·71	0·73	0·69	0·60	0·68	0·60	0·65	0·57	0·67	0·64

χ^2	10·0–11·0	11·0–12·0	12·0–13·0	13·0–14·0	14·0–15·0	15·0–16·0	16·0–17·0	17·0–18·0
No. of obs. on which mean is based	10	10	5	4	2	1	2	1
Mean $P\{T\}$	0·74	0·65	0·74	0·58	0·82	1·00	0·63	0·63

then x_k may only take values between 0 and 1 for $k = 1, 2, \ldots, s$. It is required to find the joint probability law of the product of two independent variables x and z, where x and z are as defined above. It will be noted that the elementary probability law of x will be

$$P\{x = x_j\} = p_j \quad (j = 1, 2, \ldots, s).$$

Hence when $x = x_j$ (the probability of which is p_j), the product xz will be distributed rectangularly between 0 and x_j on a proportion p_j of occasions. It follows that xz has a probability distribution which has points of discontinuity at $x_1, x_2, \ldots, x_s$, that it is distributed rectangularly between these points of discontinuity, and that

$$P\{0 < xz < x_1\} = p_1 \sum_{j=1}^{s} \frac{p_j}{x_j}, \quad P\{x_1 < xz < x_2\} = p_2 \sum_{j=2}^{s} \frac{p_j}{x_j}.$$

Generally

$$P\{x_{k-1} < xz < x_k\} = p_k \sum_{j=k}^{s} \frac{p_j}{x_j}.$$

14. If we now apply this theory to the combination of the tests of significance of T and χ^2, it is seen that we must consider the product of $P\{\chi^2\}$ and $P\{T\}$. χ^2 is a continuous variable and

$$z = \int_{\chi_0^2}^{+\infty} p(\chi^2)\, d(\chi^2) = P\{\chi^2 \geqslant \chi_0^2\} = P\{\chi^2\}$$

is distributed rectangularly between 0 and 1, and

$$x = \sum_{T=2}^{T_0} P\{T \mid r_1 r_2\} = P\{T \leqslant T_0\} = P\{T\}$$

is a discontinuous variable taking known values. The probability integral of xz is thus known from theory and $Y_{0\cdot05}$ or $Y_{0\cdot01}$ can be found to satisfy the relation

$$P\{0 < xz < Y_\epsilon\} = \epsilon.$$

These probability levels are given in Table 5. The procedure for the joint test of significance will be:

(i) calculate $P\{T\}$ as described in § 7;

(ii) calculate $P\{\chi^2\}$ in the usual way. The degrees of freedom will be the number of groups minus one;

(iii) multiply $P\{T\}$ and $P\{\chi^2\}$ together and refer to Table 5 to judge the significance of the product.

Table 5. *Values of $Y_{0\cdot05}$ and $Y_{0\cdot01}$, where $P\{P(\chi^2)\,P(T) < Y_\epsilon\} = \epsilon$*

This table may be used to judge the significance of the joint distribution of the T criterion and any other continuous criterion.

r	r_1	r_2	$Y_{0\cdot05}$	$Y_{0\cdot01}$
5	4	1	0·0312^5	0·0062^5
	3	2	0·0213	0·0043
6	5	1	0·0300	0·0060
	4	2	0·0211	0·0042
	3	3	0·0195	0·0039
7	6	1	0·0292	0·0058
	5	2	0·0197	0·0039^5
	4	3	0·0174	0·0035
8	7	1	0·0286	0·0057
	6	2	0·0188	0·0038
	5	3	0·0160	0·0032
	4	4	0·0153	0·0031
9	8	1	0·0281	0·0056
	7	2	0·0180	0·0036
	6	3	0·0153	0·0031
	5	4	0·0140	0·0028
10	9	1	0·0278	0·0056
	8	2	0·0175	0·0035
	7	3	0·0143	0·0029
	6	4	0·0143	0·0026
	5	5	0·0143	0·0025
11	10	1	0·0275$^+$	0·0055$^+$
	9	2	0·0171	0·0034
	8	3	0·0144	0·0028
	7	4	0·0144	0·0025$^+$
	6	5	0·0140	0·0024
12	11	1	0·0273	0·0055$^-$
	10	2	0·0174	0·0035$^-$
	9	3	0·0149	0·0027
	8	4	0·0142	0·0024
	7	5	0·0135	0·0022
	6	6	0·0131	0·0021
13	12	1	0·0271	0·0054
	11	2	0·0165$^+$	0·0033
	10	3	0·0151	0·0026
	9	4	0·0138	0·0023
	8	5	0·0137	0·0022
	7	6	0·0138	0·0022^5
14	13	1	0·0269	0·0054
	12	2	0·0163	0·0033
	11	3	0·0151	0·0025$^+$
	10	4	0·0135$^-$	0·0022^5
	9	5	0·0138	0·0023
	8	6	0·0136	0·0022
	7	7	0·0134	0·0022

15. The application of the joint test of significance may be illustrated by means of an example. A sample of 360 observations is available. This sample has actually been randomly drawn from a normal population of which the mean is zero and the standard deviation unity. The figures are given in Table 6. Calculations give $\chi^2 = 21{\cdot}1$ and $P\{\chi^2\} = 0{\cdot}10$. Judging by the χ^2 alone we should say probably that there is nothing out of the ordinary in the deviations of the sample from the expected values. The number of signs is 15, of which 9 are positive and 6 negative, and these are arranged in six sets. Making the appropriate calculations, we have

$$P\{6 \text{ sets} \mid 9 \text{ positive; } 6 \text{ negative}\} = \tfrac{875}{5005} = 0{\cdot}175.$$

The arrangement of signs will therefore be judged as acceptable. The joint significance of a $P\{\chi^2\} = 0{\cdot}10$ and a $P\{T\} = 0{\cdot}175$ is found, by evaluating the joint distribution, to be 0·066.

Table 6. *Sample values. Observed and expected*

Central values	−2·1 and under	−1·8	−1·5	−1·2	−0·9	−0·6	−0·3	0·0
Observation	12	10	18	26	23	42	43	49
Expectation	9·3	8·6	14·0+	21·0+	28·7	35·9	41·0+	43·0−
Deviation	+2·7	+1·4	+4·0	+5·0	−5·7	+6·1	+2·0	+6·0

Central values	+0·3	+0·6	+0·9	+1·2	+1·5	+1·8	+2·1 and over	Total
Observation	35	28	20	26	20	3	5	360
Expectation	41·0+	35·9	28·7	21·0	14·0	8·6	9·3	360
Deviation	−6·0	−7·9	−8·7	+5·0	+6·0	−5·6	−4·3	0

16. A study of the basic table (Table 1) of the function T will show that $P\{T\}$ is not a very sensitive criterion with which to judge the randomness of a sequence of signs unless the number of groups under consideration is very large. For example, if there are 10 signs, 5 of which are positive and 5 of which are negative, the probability of getting two sets of signs is 0·008. Thus the test would show, and rightly, that the chance of such an arrangement is small, but this fact would undoubtedly be recognized by a skilled computer without the use of a test at all. In the case of 10 signs the probability of three groups or less is 0·040, and this would possibly be judged non-significant. Again, let us consider an extreme case say, 10 signs, 9 of which are positive and 1 negative. The T criterion does not concern itself with the fact that the numbers 9 and 1 are exceptional, it is merely concerned with deciding whether their arrangement is exceptional given the 9 and 1. Table 1 shows that neither possible arrangement would be considered out of the ordinary. It is these points of weakness which show that the criterion T is not of great utility except in combination with χ^2. For, if we consider the 9 positive, 1 negative case, common sense tells us that the χ^2 criterion in such

a case would possibly be significant. Nine positive deviations have to be balanced by a single negative deviation, and this last is therefore likely to be big. This does not influence T; neither will the contribution of T to the joint criterion be of much weight. This is as it should be, for it is difficult to see how one can postulate a smooth alternative for 9 positive, 1 negative, two sets, and not also for 9 positive, 1 negative, three sets. Generally, however, we shall not meet such extreme cases in practice. One way of overcoming this weakness of the test would be to consider the probability of obtaining r_1 positive and r_2 negative signs together with the probability of obtaining T sets of alternate positive and negative signs given r_1 and r_2. This is simple enough when considering just a sequence of alternatives, as I have shown elsewhere, but it is not easy to fit these results to the χ^2 problem, nor, when this is possible, will the choice of a critical region be straightforward. However, the results of sampling experiments will be utilized to throw light on these points and it is hoped to discuss them, with other questions arising, in a further publication.

17. It is possible that there are other criteria, depending on the arrangement of positive and negative signs, which will be more sensitive than the T criterion chosen. For example, it is easy to calculate, given r_1 and r_2, the probability that the largest set is composed of a sequence of r' positive signs, and there are many other possibilities which might be considered. It would appear that any criterion based on sign sequences can be shown to be independent of χ^2 by means of geometrical argument, and it will be necessary therefore to consider the power of these different sign tests when referred to a specified set of alternate hypotheses.

18. The main objection to the two criteria, T and $P\{\chi^2\}.P\{T\}$, that I have proposed in this note is the one which was mentioned earlier; they are only applicable to the case where there is just one restriction on χ^2, i.e. when the totals of expected and observed frequencies have been made to agree. It is possible to work out a slightly different form of the T criterion for each additional restriction which is put on χ^2, and this has been done. It is preferable, however, to delay publication until the results of an extensive sampling experiment are complete in order to verify whether such theoretical assumptions as have been made are reasonable.

REFERENCES

Neyman, J. (1937). *Skand. Aktuar. Tidskr.* **20**, 149–99.

Neyman, J. & Pearson, E. S. (1928). *Biometrika*, **20** A, 263–94.

AN EXACT TEST FOR THE EQUALITY OF VARIANCES*

By R. L. PLACKETT, M.A.

Introduction

The problem of testing the equality of variances and covariances in normal distributions is one which has received considerable attention; we have compiled a bibliography of some sixty papers, and shall issue a survey of these in due course; only papers vital to our discussion will be considered here. A precise instance of the type of situation we are considering is as follows: measurements of height, span and tibia length are made on each of 20 Englishmen, 20 Scotsmen, 20 Welshmen and 20 Irishmen; it is required to know if the covariance matrix of the three characteristics is the same for each of the four nationalities. Nothing is known or assumed about the mean values of these characteristics in the four populations considered, nor are we interested in testing any hypothesis concerning the means, although such a hypothesis may be the object of further investigations which assume that the four covariance matrices are the same; this latter assumption is inevitably made in multivariate analysis of variance.

Wilks (1932) has already given the moments of the distribution of his criterion for testing the equality of several covariance matrices (on the hypothesis that the matrices are in fact equal) and Bishop (1939) put this criterion into an approximate workable shape. The test criterion given here differs from that of Wilks and has the advantage when one or two correlated characteristics are being measured (height or height and span, for example) that its distribution is exactly known whatever the number of populations. Nair (1939) did, it is true, give the exact distribution of the Neyman & Pearson (1931) L_1 criterion for one measured characteristic; and the exact distribution for two characteristics of Wilks's generalization of their criterion; but the form in which the distribution was obtained is very involved. It is interesting to notice that from our standpoint the problem of testing the equality of several variances (i.e. the case of one measured characteristic) is, as will appear, brought within the framework of multiple correlation theory. In the general case of more than two characteristics the moments of the distribution of our criterion, like those of Wilks, are available.

Outline of method

In the usual terminology we consider k p-variate normal distributions and are concerned with testing the hypothesis that the corresponding variances and covariances are all equal. The method we employ to test this hypothesis is essentially that which has been in use in analysis of variance since its origination by Fisher; to test the equality of a set of k quantities we test whether $(k-1)$ orthogonal linear functions of the quantities are each zero. To illustrate the application of this principle in the present instance take the particular case $p = 1$, i.e. we wish to test the equality of the variances in k univariate normal distributions. If a typical observation from the lth distribution is t_l $(l = 1, 2, \ldots, k)$, form k mutually orthogonal linear functions of the t_l such that one is

$$u = t_1 + t_2 + \ldots + t_k.$$

* Communication from the National Physical Laboratory.

If the $(k-1)$ covariances of u and each of the other linear functions are all zero then the variances of the k distributions must all be equal; this condition may be expressed by saying that the multiple correlation coefficient of u on the other linear functions is zero. Further, if there are n sample values of u then the size of sample drawn from each distribution must also be n at least, and if no observations are to be discarded the size of each sample must be n exactly. Thus, although it is not a condition of the problem that the sizes of samples drawn from the k distributions must all be equal, it is a condition of our solution.

The extension of the foregoing principle to $p > 1$ is straightforward and is considered in detail in the next section; the problem then becomes that of testing the independence of two groups of variates, the first of size p, i.e. p expressions of the form u; and the second of size $p(k-1)$ comprising all the other orthogonal linear functions. This problem has been treated by Wilks (1935, 1943) and the relevant distribution is expressible as an incomplete β-function when $p = 1$ and 2 (for all k); an exact distribution is also known when $p = 3$ and 4 for $k = 2$. Finally, since when $p = 1$ the criterion has the form of a multiple correlation coefficient, the power of the test in this instance can be calculated by virtue of the work of Fisher (1928).

Discussion of the test

A sample of n observations is drawn from each of the k p-variate normal distributions of which the lth has the covariance matrix V_{ij}^l $(l = 1, 2, \ldots, k;\ i, j = 1, 2, \ldots, p)$. It is required to test the hypothesis that

$$V_{ij}^l = V_{ij}^m \quad (l, m = 1, 2, \ldots, k). \tag{1}$$

The population means do not enter into the hypothesis and have arbitrary unknown values. Where i, j, l, m appear henceforth they will be understood to range over the values given above unless otherwise stated. The observations may be written in the form of an $n \times kp$ matrix X such that all those on the ith variate in the lth distribution are in column $(i-1)k + l$. The αth observation in this column $(\alpha = 1, 2, \ldots, n)$ is denoted by $x_{i\alpha}^l$; the order of the elements in a column is assumed to be random. If this is doubted the observations should be randomly rearranged.

We must emphasize here that the sample value of the criterion to be used to test (1) depends on this order, and there is thus, in a sense, a correspondence between $x_{i\alpha}^l$ and $x_{j\alpha}^m$, although these two quantities are, of course, uncorrelated when $l \neq m$. Most tests of a hypothesis specifying nothing about the order in which observations are made or written down are themselves independent of it; ours is not, and different computers with the same data might well come to different conclusions although this does not affect the validity of the test, the significance level being overall what it should be. There is probably some loss of power which can, however, be offset by imbuing α with a certain physical meaning; but we shall not discuss this question here. A criterion for testing normality depending on the order of arrangement of observations has been suggested by R. C. Geary (1935, pp. 316–17).

Let now
$$z_{i\alpha}^l = x_{i\alpha}^l - \left(\sum_\alpha x_{i\alpha}^l\right)\Big/ n, \tag{2}$$

and let the corresponding $n \times kp$ matrix be Z. If $G = Z'Z$, where a prime is used to denote the transpose of a matrix, then, apart from a factor n, G is the matrix of sample variances and covariances of all variables. We further define $S(k, p)$ as the sum of all (k^{2p}) signed minors

$$g_{l_1, l_2, \ldots, l_p}^{m_1, m_2, \ldots, m_p},$$

formed by rows $l_1, l_2, \ldots, l_p$ and columns $m_1, m_2, \ldots, m_p$ of G, where

$$(i-1)\,k < l_i,\ m_i \leqslant ik. \tag{3}$$

$\bar{S}(k,p)$ is similarly defined for the matrix $\bar{G} = G^{-1}$ (we shall use this notation for the inverses of matrices throughout).

We now proceed to prove the following

THEOREM: $$W(k,p) = k^{2p}/S(k,p)\,\bar{S}(k,p)$$

is distributed like Wilks's statistic for testing the hypothesis that two groups of variates of sizes p and $p(k-1)$, known to have been drawn from a (kp)-variate normal distribution, are mutually independent (Wilks, 1935, 1943). If the groups are in fact mutually independent then (1) is true.

Proof. Introduce a $k \times k$ orthogonal matrix B, the elements of whose first column are all equal (to $\pm 1/\sqrt{k}$) but which is otherwise quite arbitrary. Put

$$r = (i-1)\,k+l, \quad u = (m-1)\,p+j, \tag{4}$$

and form a $kp \times kp$ matrix A such that

$$a_{ru} = \delta_{ij}\,b_{lm}, \tag{5}$$

where $\delta_{ij} = 1$ $(i=j)$, otherwise O. Clearly A is also orthogonal. For example, suppose $k=4$, $p=2$. Apart from a factor of $\pm\frac{1}{2}$ multiplying each element, let

$$B = \begin{Vmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{Vmatrix}.$$

Then $$A = \begin{Vmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 \\ 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 \\ 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 \end{Vmatrix}.$$

When $p=1$, $A=B$. Let

$$D = XA, \quad Y = ZA, \quad C = Y'Y = A'GA. \tag{6}$$

Putting $$s = (j-1)\,k+m, \quad t = (l-1)\,p+i, \tag{7}$$

and defining $$t' = (l-1)\,p+i, \quad u' = (m-1)\,p+j \quad (l,m = 2,3,\ldots,k), \tag{8}$$

we have $$\mathscr{E}(g_{rs}) = \delta_{lm}(n-1)\,V^l_{ij}, \tag{9}$$

so that $$\mathscr{E}(c_{iu}) = (n-1)\sum_l b_{lm}\,V^l_{ij}/\sqrt{k}. \tag{10}$$

Hence when $$\mathscr{E}(c_{tu'}) = 0 \tag{11}$$

equations (1) are satisfied, because for fixed i and j equations (10) can be solved and yield

$$(n-1)\,V^l_{ij} = \mathscr{E}(c_{ij}). \tag{12}$$

Denote a typical element of the tth column of D by d_t. Then equations (11) are satisfied if and only if d_t and $d_{u'}$ are mutually independent.

A criterion for testing (11), obtained by likelihood-ratio methods, has been given by Wilks (1935, 1943). This is

$$W(k,p) = \frac{|c_{tu}|}{|c_{tj}|\,|c_{tu'}|}, \tag{13}$$

and is sometimes called the vector alienation coefficient. Let $C^{(p)}$ be the pth compound of C (Aitken, 1939, p. 90), i.e. the matrix of all $p\times p$ minors of C; and $\breve{C}^{(p)}$ the pth compound of $\breve{C} = C^{-1}$ (since $\breve{C}^{(p)}$ is the inverse of $C^{(p)}$ our notation is consistent). Then

$$W(k,p) = 1/c_{11}^{(p)}\breve{c}_{11}^{(p)} \tag{14}$$

by an application of Jacobi's theorem on the minors of the adjugate (Aitken, 1939, p. 97). Now by the Binet-Cauchy theorem (Aitken, 1939, p. 93),

$$C^{(p)} = (A')^{(p)}G^{(p)}A^{(p)}, \quad \breve{C}^{(p)} = (A')^{(p)}\breve{G}^{(p)}A^{(p)}. \tag{15}$$

Consider the elements in the first row of $(A')^{(p)}$. The first p rows of A' are of the form

$$\begin{Vmatrix} 11\ldots1 & 00\ldots0 & 00\ldots0 & \ldots & 00\ldots0 \\ 00\ldots0 & 11\ldots1 & 00\ldots0 & \ldots & 00\ldots0 \\ 00\ldots0 & 00\ldots0 & 11\ldots1 & \ldots & 00\ldots0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 00\ldots0 & 00\ldots0 & 00\ldots0 & \ldots & 11\ldots1 \end{Vmatrix}$$

apart from the factor $\pm 1/\sqrt{k}$ multiplying each element. Therefore the only non-zero elements in the first row of $(A')^{(p)}$ are those formed by taking one column from each of the p blocks of k columns into which the first p rows of A' may be divided. All the non-zero elements equal $k^{-\frac{1}{2}p}$. Then from (15)

$$c_{11}^{(p)} = S(k,p)\,k^{-p}, \quad \breve{c}_{11}^{(p)} = \breve{S}(k,p)\,k^{-p}, \tag{16}$$

so finally
$$W(k,p) = k^{2p}/S(k,p)\,\breve{S}(k,p). \tag{17}$$

This completes the proof.

Case of $p = 1$

Here
$$W(k,1) = k^2/S(k,1)\,\breve{S}(k,1),$$

where $S(k,1)$, $\breve{S}(k,1)$ are the sums of all elements of G, G^{-1} respectively. If (1) is true, $\mathbf{W}(k,1)$, the true value of $W(k,1)$, is unity. Define

$$W(k,1) = 1-R^2 \quad\text{and}\quad \mathbf{W}(k,1) = 1-\mathbf{R}^2, \tag{18}$$

so that if (1) is true, $\mathbf{R} = 0$. The distribution of $R^2 = 1-W(k,1)$ when $\mathbf{R} = 0$ is, as Wilks pointed out, well known, being that of the multiple correlation coefficient (of d_1 on $d_2, d_3, \ldots, d_k$); if in the usual notation

$$I_x(a,b) = [B(a,b)]^{-1}\int_0^x x^{a-1}(1-x)^{b-1}\,dx, \tag{19}$$

then the cumulative distribution function of $x = R^2$ is $I_x(k-1, n-k)$, values near 1 being significant; that of $x = W(k,1)$ being $I_x(n-k, k-1)$ with small values significant. Tables in convenient form have been calculated by Thompson (1941); otherwise we can convert to the variance-ratio F by

$$F = (n-k)(1-W)/(k-1)\,W. \tag{20}$$

It is clear that n must exceed k; for p variates, n exceeds pk in order that G may be non-singular.

If the matrix A is defined instead as a $kp \times kp$ orthogonal matrix, the elements of whose first column are all equal (cf. equation (5)), the problem is effectively reduced to the case $p = 1$ whatever the value of p, and we can test exactly the somewhat indefinite hypotheses

$$V^l_{i1} + V^l_{i2} + \ldots + V^l_{ip} = V^m_{j1} + V^m_{j2} + \ldots + V^m_{jp}. \tag{21}$$

This may be applied in the following manner, for take $k = p = 2$ and obtain

$$V^1_{11} + V^1_{12} = V^1_{21} + V^1_{22} = V^2_{11} + V^2_{12} = V^2_{21} + V^2_{22}. \tag{22}$$

Thus $$V^1_{11} = V^1_{22} \quad \text{and} \quad V^2_{11} = V^2_{22}. \tag{23}$$

If it is assumed $$V^1_{11} = V^2_{11}, \tag{24}$$

then $$V^1_{12} = V^2_{12}, \tag{25}$$

and conversely.

Case of $p = 2$

The distribution of $W(k, 2)$ has been given by Wilks (1935). If $x = \sqrt{[W(k, 2)]}$, the cumulative distribution function of x is

$$I_x(n-2k, 2k-2). \tag{26}$$

Small values of x are significant and n must exceed $2k$.

Case of $p \geq 3$

For $k = 2$, $p = 3$ and 4, the exact distributions are again known and have been given by Wilks in equations (35) and (37) respectively of his 1935 paper. The expressions are rather complicated and we have not reproduced them here. For other values of k and p the moments of $W(k, p)$ are available; while more recently Wald & Brookner (1941) have obtained the distribution in the form of an infinite series, calculating numerical values for the coefficients in certain instances.

For $p > 1$, (17) becomes rather intractable as a means of calculating $W(k, p)$. Indeed, for $k = 2$ and $p = 4$ it is necessary

(i) to calculate 36 sample variances and covariances,

(ii) find the inverse of an 8×8 matrix,

(iii) calculate 512 4×4 determinants,

and it is clearly better to reintroduce the matrix A in some appropriate numerical form, calculate $Y = ZA$ and $C = Y'Y$, and find $W(2, 4)$ from (13), a process which involves the evaluation of an 8×8 and two 4×4 determinants.

POWER OF THE TEST WHEN $p = 1$

From (17) the true value of R^2 is in general given by

$$1 - \mathbf{R}^2 = k^2 \Big/ \left[\left(\sum_l V^l_{11}\right)\left(\sum_l 1/V^l_{11}\right)\right], \tag{27}$$

and thus the test will have equal power for all values of the variances such that the product of their sum and the sum of their reciprocals is constant. Consequently $1 - W(k, 1)$ is distributed like the multiple-correlation coefficient in samples from a population where the

true value is given by (27). The probability density function of this distribution has been deduced by Fisher (1928) and can be integrated to give a finite series when $(n-k)$ is even.

We find easily when $k = 2$ that in the V^1_{11}, V^2_{11} quarter-plane the equipotentials are pairs of lines

$$V^1_{11} = aV^2_{11} \quad \text{and} \quad aV^1_{11} = V^2_{11}, \tag{28}$$

where

$$a = (1+\mathbf{R})/(1-\mathbf{R}). \tag{29}$$

For $k > 2$ the equipotential surfaces in k dimensions are cones through the origin situated symmetrically with regard to the co-ordinate primes.

Reverting to $k = 2$ three methods are available for testing the hypothesis that $V^1_{11} = V^2_{11}$:

(i) Fisher's z or $F = \exp(2z)$

$$= g_{11}/g_{22}. \tag{30}$$

(ii) the L_1 criterion introduced by Pearson & Neyman (1930) and later extended to $k > 2$ (Neyman & Pearson, 1931).

In the instance we are considering, i.e. equal sample sizes from both populations,

$$L_1 = 2(g_{11}g_{22})^{\frac{1}{2}}/(g_{11}+g_{22}) \tag{31}$$

$$= 2F^{\frac{1}{2}}/(1+F). \tag{31a}$$

(iii) $$W(2,1) = 4[g_{11}g_{22}-(g_{12})^2]/[(g_{11}+g_{22})^2-(2g_{12})^2].* \tag{32}$$

Thus tests (i) and (ii) are exactly equivalent, as is known, the optimum critical region being that corresponding to equal tails of the F-distribution. Criterion (iii) is that obtained by Morgan (1939) and Pitman (1939), appearing as equation (12) in Morgan's paper, to test that the variances in a normal bivarate population are equal. Morgan has compared the powers of tests (i) and (iii) for $n = 12$, 25 and 100 at a significance level of 0·10 and for these sample sizes it appears that the tests are effectively of equal power.

When n is large and, consequently, the two populations being independent, $(g_{12})^2/g_{11}g_{22}$ is converging in probability to zero,

$$W(2,1) \sim L_1^2. \tag{33}$$

The cumulative distribution functions of criteria (ii) and (iii) are respectively $I_x\left(\frac{n-1}{2}, \frac{1}{2}\right)$ (Nayer, 1936) ($x = L_1^2$) and $I_x\left(\frac{n-2}{2}, \frac{1}{2}\right)$ ($x = W(2,1)$). Generally, $W(k,1)$ for large n converges in probability to the harmonic mean of the sample variances divided by their arithmetic mean; L_1 (for equal sample sizes) is exactly equal to the geometric mean divided by the arithmetic mean.

Example of the use of the test for a case with $k=4$, $p=1$

It is not easy to calculate $W(k,1)$ from equation (17) if $k > 3$. Indeed, the main value of (17) lies in showing the form of solution, and in establishing that this is independent of the particular orthogonal transformations used. In the following example, therefore, orthogonal transformations are made at once and the multiple correlation coefficient is calculated from the numerical data. This procedure is far quicker than that involved in calculating $W(4,1)$ from (17).

* See Appendix.

Below are given samples of 10 from each of four univariate normal populations:

x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4
−20	+24	+ 4	+52	+ 7	+15	+ 8	− 8
− 1	+18	+ 9	−24	+ 5	+24	− 1	+56
−11	+27	−27	0	+18	−12	+ 1	−64
+10	+21	+ 5	+48	+13	−24	− 4	+12
− 4	−48	− 3	+48	− 6	+12	+ 5	−12

which have mean zero and standard deviations respectively 10, 30, 10, 40. Make the following orthogonal transformation:

$$y_1 = x_1+x_2+x_3+x_4, \quad y_2 = x_1+x_2-x_3-x_4,$$
$$y_3 = x_1-x_2+x_3-x_4, \quad y_4 = x_1-x_2-x_3+x_4,$$

and obtain

y_1	y_2	y_3	y_4	y_1	y_2	y_3	y_4
+60	−52	−92	+ 4	+22	+22	+ 8	−24
+ 2	+32	+14	−52	+84	−26	−76	+38
−11	+43	−65	−11	−57	+69	+95	−57
+84	−22	−54	+32	− 3	−19	+21	+53
− 7	−97	− 7	+95	− 1	+13	− 1	−35

Form the matrix of sums of squares and cross-products, i.e. C. This is

$$\begin{Vmatrix} +18636\cdot1 & -9646\cdot9 & -18232\cdot9 & +7325\cdot1 \\ -9646\cdot9 & +21784\cdot1 & +12018\cdot1 & -19015\cdot9 \\ -18232\cdot9 & +12018\cdot1 & +28692\cdot1 & -9445\cdot9 \\ +7325\cdot1 & -19015\cdot9 & -9445\cdot9 & +22008\cdot1 \end{Vmatrix}.$$

The matrix of sample correlation coefficients is therefore

$$\begin{Vmatrix} 1 & -0\cdot4788 & -0\cdot7885 & +0\cdot3617 \\ -0\cdot4788 & 1 & +0\cdot4807 & -0\cdot8685 \\ -0\cdot7885 & +0\cdot4807 & 1 & -0\cdot3759 \\ +0\cdot3617 & -0\cdot8685 & -0\cdot3759 & 1 \end{Vmatrix}.$$

Hence the multiple correlation coefficient of y_1 on y_2, y_3, y_4 is given by $R^2 = 0\cdot$
by the approximation indicated in the last paragraph of the preceding sectic
the true value obtained from equation (27) with variances in the ratio 1 :
The upper 10 and 5 % levels of significance, obtained from Thompson's tables with $\nu_1 = n-k = 6$, $\nu_2 = k-1 = 3$, are respectively 0·622 and 0·704. We find $L_1 = 0\cdot565$, the 5 and 1 % levels obtained from Nayer's (1936) tables being respectively 0·797 and 0·719, so that this test gives a more significant result than the one based on R^2. The relative merits of L_1 and the test we have provided, which cannot be judged on the results of one example, remain a problem to be investigated.

* I.e. calculated from 1 − (harmonic mean of g_{tt})/(arithmetic mean of g_{tt}).

Summary

An exact test has been put forward for the equality of variances and covariances in any number k of 1- or 2-variate normal populations; the test is also exact for two 3- or 4-variate populations; but is restricted in application to equal sample sizes n from the k populations where n exceeds pk, p being the number of variates. The moments of the criterion are available for k p-variate populations where the statistic used is equivalent to that employed by Wilks (1935) to test the independence of two groups of variates (of sizes p and $p(k-1)$), and has the same distribution. In the univariate case the power of the test is known as a function of one parameter. Comparison with the L_1 criterion has already been made when $p = 1$ and $k = 2$, the tests being practically the same, and an example worked out of the use of the test when $p = 1$.

Our thanks are due to E. C. Fieller for drawing our attention to the papers by Morgan and Pitman and suggesting that the test given there for the equality of two variances might be extended to more than two; also to Prof. E. S. Pearson for pointing out the need of certain explanatory additions.

The work described above has been carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory.

REFERENCES

Aitken, A. C. (1939). *Determinants and Matrices.* Oliver and Boyd, Edinburgh.

Bishop, D. J. (1939). On a comprehensive test of the homogeneity of variances and covariances in multivariate problems. *Biometrika*, **31**, 31.

Fisher, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. *Proc. Roy. Soc.* A, **121**, 654.

Geary, R. C. (1935). The ratio of the mean deviation to the standard deviation as a test of normality. *Biometrika*, **27**, 310.

Morgan, W. A. (1939). A test for the significance of the difference between the two variances in a sample from a normal bivariate population. *Biometrika*, **31**, 13.

Nair, U. S. (1939). The application of the moment function in the study of distribution laws in statistics. *Biometrika*, **30**, 274.

Nayer, P. P. N. (1936). An investigation into the application of Neyman and Pearson's L_1 test, with tables of percentage limits. *Statist. Res. Mem.* **1**, 38.

Neyman, J. & Pearson, E. S. (1931). On the problem of k samples. *Bull. int. Acad. Cracovie*, Série A, p. 460.

Pearson, E. S. & Neyman, J. (1930). On the problem of two samples. *Bull. int. Acad. Cracovie*, Série A, p. 73.

Pitman, E. J. G. (1939). A note on normal correlation. *Biometrika*, **31**, 9.

Thompson, C. M. (1941). Tables of percentage points of the incomplete beta-function. *Biometrika*, **32**, 168.

Wald, A. & Brookner, R. J. (1941). On the distribution of Wilks' statistic for testing the independence of several groups of variates. *Ann. Math. Statist.* **12**, 137.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, **24**, 471.

Wilks, S. S. (1935). On the independence of k sets of normally distributed statistical variables. *Econometrika*, **3**, 309.

Wilks, S. S. (1943). *Mathematical Statistics.* Princeton University Press.

APPENDIX

As an illustration of the algebraic form of $W(k, 1)$ the Editor has suggested to me that it might be helpful to show the relation of the general formula (17) to the matrix G in this simple case when $k = 2$. Here, using a common notation for a sample mean

$$g_{11} = \sum_{\alpha=1}^{n} (x^1_{1\alpha} - x^1_{1\cdot})^2, \quad g_{22} = \sum_{\alpha=1}^{n} (x^2_{1\alpha} - x^2_{1\cdot})^2, \quad g_{12} = \sum_{\alpha=1}^{n} (x^1_{1\alpha} - x^1_{1\cdot})(x^2_{1\alpha} - x^2_{1\cdot}),$$

$$G = \begin{bmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{bmatrix}, \quad \breve{G} = \begin{bmatrix} g_{22} & -g_{12} \\ -g_{12} & g_{11} \end{bmatrix} \Big/ \{g_{11}g_{22} - g^2_{12}\},$$

$$S(2, 1) = g_{11} + g_{22} + 2g_{12}, \quad \breve{S}(2, 1) = \frac{g_{11} + g_{22} - 2g_{12}}{g_{11}g_{22} - (g_{12})^2}.$$

Whence, using (17), (32) is at once obtained for $W(2, 1)$. For $k > 2$ the full expression for $\breve{S}(k, 1)$ in terms of the g's is complicated and the matrix notation becomes essential.

THE ESTIMATION FROM INDIVIDUAL RECORDS OF THE RELATIONSHIP BETWEEN DOSE AND QUANTAL RESPONSE

By D. J. FINNEY

Lecturer in the Design and Analysis of Scientific Experiment, University of Oxford

1. Introduction

A type of biometric problem frequently encountered by the statistician is that which requires the estimation and study of a relationship between dose and response. 'Dose' is here a general term indicating the magnitude of a stimulus applied to certain test subjects, and 'response' is a measure of the effect which the stimulus produces on the subjects. When the test subjects are living matter, whether plants, animals or bacteria, pieces of tissue or single cells, the response to a specified dose is unlikely to be constant in repeated trials, and regression methods must be used in the estimation of the relationship.

In some classes of data, the response is 'all-or-nothing' or quantal, and cannot be measured quantitatively. Ordinary regression methods are then no longer applicable; methods based on the transformation of the proportion of subjects showing the response at any dose level to the normal equivalent deviate (Gaddum, 1933), or to the probit (Bliss, 1934*a*, *b*), however, have proved very powerful for simplifying the statistical analysis. In recent years, full accounts of the underlying theory of these transformations, and of their application, have been published by various authors (see, for example, Bliss, 1935*a*, *b*; Finney, 1947, 1948). An additional difficulty sometimes found is that the intensity of the stimulus cannot be *selected* in advance of a test, but can only be *measured* after the test has taken place; only rarely will two or more subjects happen to receive exactly the same dose, and more usually the records consist of a list of doses with, for each, a statement of whether a single subject receiving that dose responded or not.* For example, in some methods for the testing of insecticidal potency, poison bait is offered to individual insects; the dose received by any insect cannot be specified in advance, and must instead be measured as the amount of poison ingested.

Data from experiments of this kind do not give empirical values for the proportion of subjects responding at each dose level, except in the trivial sense that every dose shows either zero or 100 % responding. Nevertheless, as Bliss (1938) has pointed out, the probit method can still be applied to estimation of the dose-response relationship. He has given a numerical example, though without showing full details of the working, but has admitted that assessment of the error of estimation presents some theoretical difficulties (Finney, 1947, § 43). An interesting example of experimental results requiring this type of analysis has recently been brought to the notice of the writer by Mr R. W. Gilliatt. These introduce an additional complication, since the dose is expressed in terms of two measurements, and a probit plane (Finney, 1943) or other bivariate regression function must therefore be estimated. An account of the analysis, with computational details, may help those who have encountered analogous problems in biological or other investigations.

* When response does not involve death or serious alteration of the test subject, one subject may be used many times; the example discussed in this paper is an instance. The form of the data will be the same, though the interpretation may require that tolerance variation between and within subjects be distinguished.

2. The data

Research in human physiology has demonstrated that, under carefully controlled experimental conditions, a transient reflex vaso-constriction in the skin of the digits may follow a single deep breath (Bolton, Carmichael & Stürup, 1936). Gilliatt (1947) has found that the response depends in part on the volume of air taken in by the subject. Plethysmographic measurement of the volume changes in a finger was used to indicate the occurrence of a response, but assessment of the degree of vaso-constriction, in order to relate this to the inspiratory stimulus, was not practicable. Thus the records obtained for each test show only

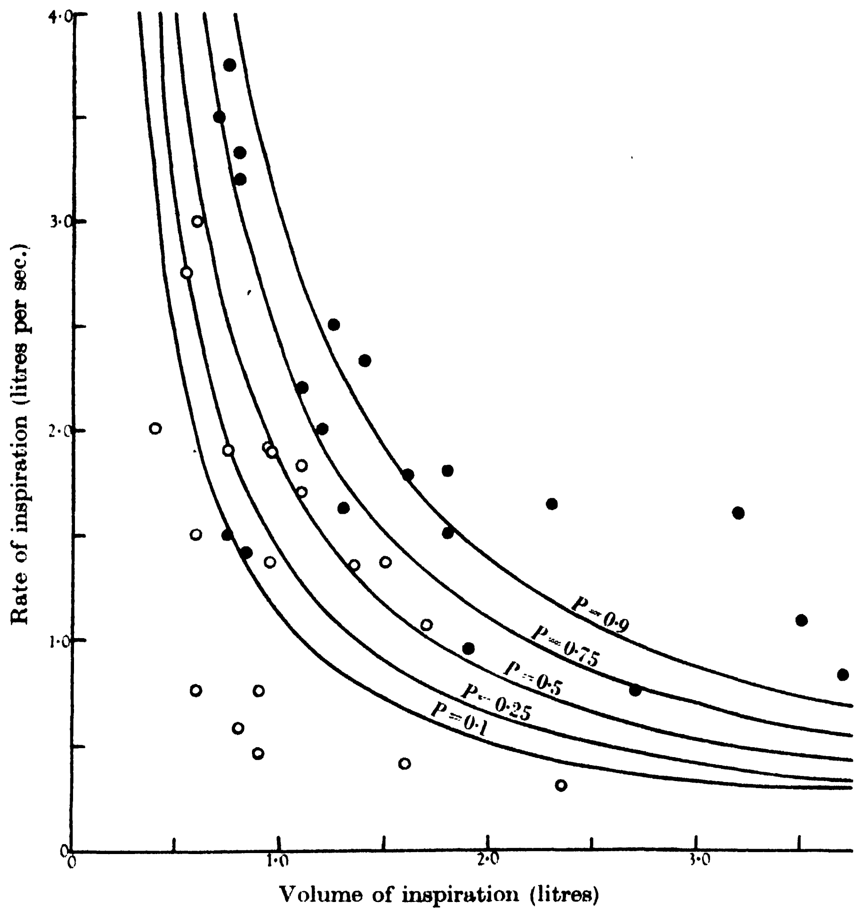


Fig. 1. Contours of dose-response surface for 0·1, 0·25, 0·5, 0·75 and 0·9 frequency of response, estimated from three-parameter equation. ○ no vaso-constriction; ● vaso-constriction.

the volume of air inspired, the average rate of inspiration, and whether or not vaso-constriction was produced. The above brief outline is sufficient for appreciation of the statistical problem, but a full account of the experimental procedure may be found in Gilliatt's paper; the results discussed here are presented in his Fig. 5.

The data, which Mr Gilliatt has kindly made available to the writer, were obtained from thirty-nine tests, in twenty of which vaso-constriction occurred. Tests were made on three different subjects, nine on D.W., eight on V.P.W., and twenty-two on S.J.S.; the results of the tests, with the subjects in this order, are shown in Table 1. In Fig. 1 are shown the thirty-nine combinations of volume in litres (V) and rate of inspiration in litres per second

Table 1. *Experimental data and details of calculations*

Volume in litres (V)	Rate in litres per sec. (R)	x_1	x_2	Response	Y	w	y	wx_1	wx_2	wy	P
3·7	0·825	1·57	0·92	+	6·8	0·18	7·26	0·2826	0·1656	1·3068	0·962
3·5	1·09	1·54	1·04	+	7·3	0·08	7·68	0·1232	0·0832	0·6144	0·989
1·25	2·5	1·10	1·40	+	6·5	0·27	7·02	0·2970	0·3780	1·8954	0·931
0·75	1·5	0·88	1·18	+	3·7	0·34	8·97	0·2992	0·4012	3·0498	0·098
0·8	3·2	0·90	1·51	+	5·8	0·50	6·53	0·4500	0·7550	3·2650	0·789
0·7	3·5	0·85	1·54	+	5·6	0·56	6·42	0·4760	0·8624	3·5952	0·741
0·6	0·75	0·78	0·88	−	1·3	0·00	1·05	0·0000	0·0000	0·0000	0·000
1·1	1·7	1·04	1·23	−	5·1	0·63	3·74	0·6552	0·7749	2·3562	0·530
0·9	0·75	0·95	0·88	−	2·4	0·04	2·06	0·0380	0·0352	0·0824	0·005
0·9	0·45	0·95	0·65	−	1·0	0·00	0·76	0·0000	0·0000	0·0000	0·000
0·8	0·57	0·90	0·76	−	1·4	0·00	1·14	0·0000	0·0000	0·0000	0·000
0·55	2·75	0·74	1·44	−	4·3	0·53	3·53	0·3922	0·7632	1·8709	0·248
0·6	3·0	0·78	1·48	−	4·8	0·63	3·72	0·4914	0·9324	2·3436	0·430
1·4	2·33	1·15	1·37	+	6·6	0·24	7·09	0·2760	0·3288	1·7016	0·950
0·75	3·75	0·88	1·57	+	6·0	0·44	6·66	0·3872	0·6908	2·9304	0·848
2·3	1·64	1·36	1·21	+	7·1	0·11	7·51	0·1496	0·1331	0·8261	0·982
3·2	1·6	1·51	1·20	+	8·0	0·01	8·30	0·0151	0·0120	0·0830	0·999
0·85	1·415	0·93	1·15	+	3·9	0·40	7·87	0·3720	0·4600	3·1480	0·128
1·7	1·06	1·23	1·03	−	5·2	0·63	3·72	0·7749	0·6489	2·3436	0·563
1·8	1·8	1·26	1·26	+	6·7	0·21	7·17	0·2646	0·2646	1·5057	0·958
0·4	2·0	0·60	1·30	−	2·6	0·06	2·23	0·0360	0·0780	0·1338	0·007
0·95	1·36	0·98	1·13	−	4·1	0·47	3·41	0·4606	0·5311	1·6027	0·179
1·35	1·35	1·13	1·13	−	5·1	0·63	3·74	0·7119	0·7119	2·3562	0·533
1·5	1·36	1·18	1·13	−	5·4	0·60	3·62	0·7080	0·6780	2·1720	0·662
1·6	1·78	1·20	1·25	+	6·3	0·34	6·86	0·4080	0·4250	2·3324	0·897
0·6	1·5	0·78	1·18	−	3·0	0·13	2·58	0·1014	0·1534	0·3354	0·025
1·8	1·5	1·26	1·18	+	6·2	0·37	6·79	0·4662	0·4366	2·5123	0·894
0·95	1·9	0·98	1·28	−	5·0	0·64	3·75	0·6272	0·8192	2·4000	0·489
1·9	0·95	1·28	0·98	+	5·2	0·63	6·28	0·8064	0·6174	3·9564	0·577
1·6	0·4	1·20	0·60	−	2·4	0·04	2·06	0·0480	0·0240	0·0824	0·005
2·7	0·75	1·43	0·88	+	5·6	0·56	6·42	0·8008	0·4928	3·5952	0·727
2·35	0·03	1·37	0·48	−	2·8	0·09	2·41	0·1233	0·0432	0·2169	0·015
1·1	1·83	1·04	1·26	−	5·2	0·63	3·72	0·6552	0·7938	2·3436	0·600
1·1	2·2	1·04	1·34	+	5·7	0·53	6·47	0·5512	0·7102	3·4291	0·767
1·2	2·0	1·08	1·30	+	5·8	0·50	6·53	0·5400	0·6500	3·2650	0·776
0·8	3·33	0·90	1·52	+	5·8	0·50	6·53	0·4500	0·7600	3·2650	0·806
0·95	1·9	0·98	1·28	−	5·0	0·64	3·75	0·6272	0·8192	2·4000	0·489
0·75	1·9	0·88	1·28	−	4·3	0·53	3·53	0·4664	0·6784	1·8709	0·243
1·3	1·625	1·11	1·21	+	5·4	0·60	6·34	0·6660	0·7260	3·8040	0·665
					Totals	14·29		14·9980	17·8375	74·9914	
					Means			1·0495	1·2483	5·2478	

Swx_1^2	Swx_1x_2	Swx_2^2	Swx_1y	Swx_2y	Swy^2
16·235606	18·338532	22·783383	79·738989	94·125032	433·586326
15·741078	18·721261	22·265669	78·706859	93·608054	393·541643
0·494528	−0·382729	0·517714	1·032130	0·516978	40·044683

(R), together with indications of whether or not the subject responded under these conditions. Inspection of Fig. 1 shows that, in general, when both V and R were small no response occurred, when either was large (unless the other was very small) the response occurred, and in an intermediate region the proportion of responses increased as either V or R increased. There was no sharply defined threshold separating combinations of V and R giving the response from those giving no response; instead, there appeared to be a probability of response ranging from practical certainty under some conditions to zero under others.

As an aid to fuller understanding of the influence of breathing on vaso-constriction, examination of the relationship between V, R, and the probability of response seemed desirable. Since so few observations were available for each subject, the data were unlikely to be sufficient to show differences between subjects; this point is discussed later, but in the main analysis the distinction between subjects is ignored. For any form of response assessment, the testing of one subject many times must introduce a danger that the result of one test will be affected not only by its own stimulus but by preceding stimuli and by the effects they produced. In this investigation, each subject was given a number of preliminary tests until he appeared to have settled into the routine. The observations recorded in Table 1 were obtained after these preliminary trials; they are tabulated in the order of testing, and show no indication of effects of previous history, but clearly such effects would have to be very pronounced if they were to be detectable on this amount of data.

3. METHOD OF ANALYSIS

Preliminary examination of the data suggested that the occurrence of a response was largely determined by the magnitude of VR, the product of volume and rate, curves on which the probability of response has a constant value being approximately hyperbolae of the form

$$VR = \text{constant}. \tag{1}$$

A little consideration shows that an equation of this type is more reasonable than an equation linear in V and R, though the data are almost certainly inadequate for discriminating between many alternative types of relationship that might be postulated. A system of curves similar to, but rather more general than, equation (1), namely,

$$V^{\beta_1} R^{\beta_2} = \text{constant}, \tag{2}$$

was selected for trial; this equation may alternatively be regarded as representing a series of parallel linear relationships

$$\beta_1 \log V + \beta_2 \log R = \text{constant} \tag{2a}$$

between the logarithms of volume and rate for a fixed probability of response.

A specified combination of V and R will not necessarily always give the same result (response or no response) with a subject, for, even though the subject is unaltered, minor uncontrolled variations in his environment may affect his susceptibility to the applied stimulus. For a particular value of V, the threshold value of R (the value which under the conditions prevailing at any instant would be just sufficient to produce a response) will have a frequency distribution; similarly, for a particular R, there will be a frequency distribution of threshold values of V. If these distributions may be taken as normal in $\log V$ and $\log R$, and, for simplicity, they are supposed to be such that the mean of either logarithm is linearly related to the selected value of the other, then the probability of response will be determined by an expression of the form

$$\beta_1 \log V + \beta_2 \log R,$$

and the threshold values of this quantity will be normally distributed. If x_1 and x_2 are written for $\log(10V)$ and $\log(10R)$ respectively (the factor of 10 is introduced in order to make x_1 and x_2 always positive), this statement enables the probability of response, P, to be expressed as

$$P = \int_{-\infty}^{\alpha+\beta_1x_1+\beta_2x_2} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}u^2} du, \tag{3}$$

where α, β_1 and β_2 are parameters to be estimated from the data. The estimation may be regarded as the fitting of a probit regression plane, for Y, the probit of P being given by

$$Y = 5+\alpha+\beta_1x_1+\beta_2x_2. \tag{4}$$

Substitution of the value of Y corresponding to a specified probability gives the required linear relationship, equation (2*a*), between x_1 and x_2 for that probability, from which the estimated curves of constant probability, equation (2), may easily be derived.

The procedure for fitting a probit plane has been described elsewhere (Finney, 1943, 1947, § 31), and its chief features need no alteration for application to individual records. Providing that a first approximation to the equation can be guessed, repeated cycles of computation will give values for the parameters which approach more and more closely to the maximum likelihood estimates. Care in the choice of the first approximation will reduce the number of cycles needed; a poor choice will delay the convergence, though it will not affect the ultimate result. Since only a single observation is available for each combination of x_1 and x_2, every working probit is either a maximum or minimum value, according to whether or not the response occurs. When there is only one dose factor, in the fitting of a probit regression line to records of individuals, grouping of doses and treatment of the observations in a group as if they related to an average dose may reduce the labour of the early computing cycles, but, since it will tend to give an underestimate of the regression coefficient, the final cycle may need to use the detailed records. Bliss (1938) has given an example illustrating grouping of this kind. Grouping is less easily applied, however, when two or more dose factors have to be used, and, for the data under discussion, the individual records were used throughout except in the formation of the first approximation.

In the standard form of probit analysis, with moderately large numbers of observations at each level of dose, a χ^2 is usually computed for testing the significance of discrepancies between the data and the fitted equation; this χ^2 is numerically the same as would be obtained by calculation from expected and observed numbers of responses and non-responses for each dose. If there are few observations in any dose group, the expected number of responses or of non-responses (or of both) is likely to be small, and, as is well known, χ^2 may then fail to follow the sampling distribution tabulated for that statistic. Data of the type under discussion here are extreme examples of this situation, the number of observations for each dose being reduced to unity, so that any disturbance of the χ^2 distribution is likely to be encountered in its most acute form. No complete theoretical investigation of this matter has yet been made, but the practical implications are discussed more fully in § 5.

On the assumption that the estimate of equation (4) is an adequate representation of the data, lines of constant response probability may be obtained for any specified probability; these may be plotted according to equation (2) on a V, R scale. Standard statistical processes also enable fiducial limits to be assigned to the position of any of these curves. The difficulty of dealing with the estimation of error for individual records, and the inadequacy of the data for any sensitive test of whether equation (2) is a satisfactory representation of the

system of curves, throw doubts on the exact interpretation of these fiducial limits. Nevertheless, they give some idea of the confidence that can be attached to the estimated curves, at least for moderate values of V and R; for extremes of either measurement, far more extensive data would be needed before much faith could be placed in the fitted equation.

4. COMPUTATIONS FOR ESTIMATING THE THREE-PARAMETER EQUATION

In this and the two succeeding sections, the computations for Gilliatt's data will be described in detail. The first five columns of Table 1 show the thirty-nine pairs of values of V and R which occurred in the experiments, followed by the corresponding values of x_1 and x_2, together with a statement of whether or not the subject responded. Before the probit computations could be initiated, a first approximation to equation (4) was needed; this was obtained with the aid of the suggestion, from the plotting of the data shown in Fig. 1, that the constant probability curves were approximately the hyperbolae of equation (1), or alternatively

$$x_1 + x_2 = \text{constant}.$$

As Bliss (1938) has pointed out, there is no objection to the use of overlapping groups in the formation of the first approximation. The data were therefore grouped according to the value of $(x_1 + x_2)$, as shown below, and the proportion of responses in each group was obtained from Table 1:

$x_1 + x_2$	Responses	Proportion (p)	Probit of p	First approximation
1·5–1·9	0/7	0·00	—	3·3
1·6–2·0	0/7	0·00	—	3·6
1·7–2·1	2/7	0·29	4·4	3·9
1·8–2·2	2/9	0·22	4·2	4·2
1·9–2·3	3/14	0·21	4·2	4·5
2·0–2·4	8/19	0·42	4·8	4·8
2·1–2·5	13/24	0·54	5·1	5·1
2·2–2·6	17/25	0·68	5·5	5·4
2·3–2·7	16/17	0·94	6·6	5·7
2·4–2·8	12/12	1·00	—	6·0

Each proportion was regarded as an estimate for the median value of $(x_1 + x_2)$ in the group, i.e. 1·7, 1·8, 1·9, ..., and its probit was read from one of the standard tables (Finney, 1947, Table I; Fisher & Yates, 1947, Table IX). As may be seen above, these probits were fairly well fitted by the guessed equation

$$Y = -1{\cdot}8 + 3(x_1 + x_2), \tag{5}$$

which was therefore used as a first approximation to equation (4).

A first set of expected probits was calculated from equation (5), and inserted as Y in an earlier version of Table 1. A cycle of routine probit calculations, just as described in the next two paragraphs, then led to an improved approximation to the required estimate, on which a second cycle of improvement was based. The figures shown in Table 1 relate to the fourth of these cycles, based upon the approximation

$$Y = -9{\cdot}127 + 6{\cdot}666x_1 + 5{\cdot}906x_2 \tag{6}$$

from the third cycle. Equation (6) is very different from equation (5), suggesting that more care might have been given to the selection of a first approximation; that the grouping

adopted would lead to underestimation of the regression coefficients was expected, but insufficient allowance for this was made. Of course the 'improvement' in the approximations refers to their approach to the solution of the maximum likelihood equations, and is not necessarily always an approach to the true relationship.

The column of expected probits, Y, in Table 1 was calculated by substitution of pairs of values x_1, x_2 in equation (6); one decimal place here is quite sufficient. The weighting coefficient, w, for each observation was then read from tables (Finney, 1947, Table II; Fisher & Yates, 1947, Table XI) and entered in its column. The working probit, y, takes a maximum value for every observation giving a response and a minimum value for every observation giving no response, since these give empirical rates of 100 % and zero respectively; values of y were read directly from Finney's table (1947, Table III; or, less simply for the minimum values, from Fisher & Yates, 1947, Table XI). The numbers of decimal places shown for the entries in Table 1 are sufficient for data of this type; indeed possibly one decimal for w and for y would be enough. Columns wx_1, wx_2, and wy were then filled, and the weighted sums of squares and products of deviations, required for the calculation of the regression of y on x_1 and x_2, were completed at the bottom of the table.

The equations giving the estimates of the regression coefficients, b_1 and b_2, are

$$0{\cdot}494528b_1 - 0{\cdot}382729b_2 = 1{\cdot}032130,$$
$$-0{\cdot}382729b_1 + 0{\cdot}517714b_2 = 0{\cdot}516978.$$

Later calculations use the variances and covariance of b_1 and b_2; the equations were therefore solved by first obtaining the matrix inverse to that formed by the coefficients of b_1 and b_2 (Finney, 1943, 1947, § 31; Fisher, 1946, § 29). This matrix is

$$V = \begin{pmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{pmatrix} = \begin{pmatrix} 4{\cdot}726144 & 3{\cdot}493883 \\ 3{\cdot}493883 & 4{\cdot}514482 \end{pmatrix}; \tag{7}$$

the accuracy of the data is insufficient to need the number of decimal figures shown here, but their retention assists the checking and maintains the internal consistency of the analysis. Now

$$b_1 = 1{\cdot}032130v_{11} + 0{\cdot}516978v_{12}$$
$$= 6{\cdot}68426,$$

and similarly

$$b_2 = 5{\cdot}94003.$$

The estimate of equation (4) is then

$$Y = \bar{y} + b_1(x_1 - \bar{x}_1) + b_2(x_2 - \bar{x}_2)$$

or

$$Y = -9{\cdot}182 + 6{\cdot}6843x_1 + 5{\cdot}9400x_2, \tag{8}$$

a result which differs little from equation (6) and may be regarded as a sufficiently close approximation to the maximum likelihood estimate. Since

$$b_2/b_1 = 0{\cdot}889, \tag{9}$$

equation (8) may be transformed to give

$$VR^{0{\cdot}889} = \text{constant} \tag{10}$$

as the relationship estimated to exist between V and R for a specified probability; the value of the constant can be obtained by substitution of the probit of the probability in equation (8), a process which gives 1·10, 1·36, 1·71, 2·16 and 2·66 for probabilities of 10, 25, 50, 75 and 90 % respectively. Typical contours have been drawn in Fig. 1 so as to indicate the form of the relationship.

5. GOODNESS OF FIT

When probit analysis is applied to data containing many observations in each dose group, the weighted sum of squares of deviations between the empirical probits and the predictions from the fitted equations is a χ^2, with degrees of freedom equal to the number of dose groups reduced by the number of fitted parameters. If S_{uv} is written for the weighted sum of products of deviations of variates u and v, application of this method here would give

$$\begin{aligned}\chi^2_{[36]} &= S_{yy} - b_1 S_{x_1 y} - b_2 S_{x_2 y} \\ &= 40{\cdot}045 - 6{\cdot}6843 \times 1{\cdot}0321 - 5{\cdot}9400 \times 0{\cdot}5170 \\ &= 30{\cdot}08. \qquad (11)\end{aligned}$$

When the dose groups are small, however, the χ^2 so calculated cannot be trusted as an indicator of the significance of deviations from the fitted equation, and it is presumably most unreliable when each group is reduced to a single observation. Apart from slight discrepancies caused by imperfect approximation to the maximum likelihood solution, the χ^2 in equation (11) is algebraically identical with that which would be derived, by the usual form of calculations, from comparison of observed numbers responding and not responding in each group with expectations computed from the fitted equation. As is well known from the study of contingency tables, when the expectations in some classes are small the sampling distribution of such a χ^2 may be very different from that shown in the standard tables (Finney, 1947, Table VI; Fisher & Yates, 1947, Table IV); with data from individual records, no class can have an expectation greater than unity, and for many the expectation will be very much less, so that the discrepancy from the tabulated χ^2 distribution is likely to be serious.

The general effect of small expectations on the random sampling distribution of χ^2 appears to be that the mean value remains about equal to the number of degrees of freedom, but that the variance in repeated sampling is increased. Consequently, samples from a population according with the null hypothesis are likely to show an excess of very high and very low values, as judged by the tables of χ^2. Thus there is little danger that significant evidence of deviations from expectation will be overlooked in an uncritical application of the test, though apparently significant values of χ^2 need to be examined with care before they are regarded as evidence sufficient to justify rejection of the null hypothesis. Low values, as in Gilliatt's data 30 with 36 degrees of freedom, need cause little alarm, for they clearly indicate no serious deviation from expectation. High values may in the first instance be compared with the standard tables of the χ^2 distribution; if they fall beyond the significance level, a closer examination should be made before judging the null hypothesis to be untenable, for the apparent significance may be due to large contributions from one or two aberrant points. Gilliatt's data provide an illustration of this. The expected probits for each pair of values of x_1 and x_2 in Table 1 have been calculated from equation (8), and the probabilities, P $(= 1 - Q)$, corresponding to these have been entered in the last column of the table; P is then the expectation of the number of responses for each dose. The χ^2 obtained from the observed and expected numbers in seventy-eight classes is easily seen to be the sum of Q/P for all doses giving a response, plus P/Q for all giving no response. Inspection of the column for P shows small contributions to χ^2 everywhere, except for two instances of responses with probabilities of only 0·098 and 0·128, contributing 9 and 7 respectively; clearly the occurrence of these two responses as the most extreme events in thirty-nine trials need not be regarded

as serious evidence against the null hypothesis. The result of calculating χ^2 by this more laborious process is a total of 30·3, which agrees closely with that already given in equation (11).

One method of modifying a χ^2 test so as to remove its extreme sensitivity to deviations from small expectations is to combine expected and observed frequencies over several adjacent groups, so as to obtain groups with larger expectations; the number of degrees of freedom is then taken as the number of remaining groups less the number of fitted parameters. Of course the groups must be chosen objectively, and without regard to the agreement between the frequencies. The statistic still will not follow the χ^2 distribution exactly, but the approximation should be fairly satisfactory under the usual restriction that the groups be so chosen that none of the expected frequencies is small. This procedure often has to be adopted in probit analysis because of small expectations at very low or very high doses (Finney, 1947, § 18). With individual records, however, only very extensive grouping will give expectations sufficiently large for the χ^2 test to be trusted; the reduction of a large χ^2 to a value below the significance level might then appear indicative of an insensitive test rather than of absence of serious discrepancies.*

Probably no completely satisfactory solution of the difficulty is to be expected. Individual records usually arise from experimental work in which the obtaining of large numbers of observations presents considerable difficulty. Often the whole series will consist of less than fifty observations, and, unless previous information enables the range of doses to be chosen satisfactorily, many of the observations will be made at doses for which response is either almost certain or almost impossible. Even if the individual dose-tolerances could be measured directly, a test of normality of their distribution (which is what the χ^2 test attempts to provide) could not be very sensitive when based on only fifty measurements; if, instead, only quantal data are available, indicating merely whether a dose is below or above the tolerance value, a sensitive normality test is still less likely to exist (Finney, 1947, § 43).

Gilliatt's data, a series of only thirty-nine observations, provide an extreme instance of the difficulty of formulating a sensitive test of goodness of fit. Nevertheless, an attempt has been made to examine the discrepancies between the observations and the null hypothesis expressed by equation (4). In Table 2 are compared the observed and expected frequencies when the data are grouped according to the value of $VR^{0\cdot889}$. This is equivalent to a grouping based on the value of Y, the expected probit in equation (8), and, as this quantity had been evaluated for each observation in order to give P, it was used in the construction of Table 2. Since three parameters have been estimated from the data, four groups is the least number for giving a χ^2 test. The limits of the groups were chosen so as to give similar numbers of observations in each. Inspection of Table 2 shows that the groups are still too small for a χ^2 test to be trusted, thus suggesting that the data are inadequate for any useful test of goodness of fit to be made. The only anomaly in Table 2 is the occurrence of two responses where the expectation is 0·3, and this is clearly insufficient to cause much worry.

The inadequacy of the data for detecting any differences in sensitivity between the three subjects may be seen from Table 3. The first nine entries in Table 1 relate to D.W., and sum-

* In his discussion of the analysis of individual records, Bliss (1938) suggests adjustment of the χ^2 test, not by altering the calculation of the statistic but by reducing the number of degrees of freedom allotted to it; he gives an empirical rule for the reduction, based upon the expectations in terminal dose groups. This method, however, not only lacks any theoretical basis, but seems liable to have an effect opposite to that which is needed; it will attribute significance to high values of χ^2 even more readily than will the unadjusted test.

mation of the values of P gives the expected number of responses for this subject; similarly the next eight and the last twenty-two entries give the numbers for V.P.W. and S.J.S. respectively. Inspection of Table 1 shows that the tests on each subject were fairly widely distributed over the range of values of x_1 and x_2. Table 3 shows excellent agreement between totals of observed and expected responses for each subject, thus suggesting that any individual differences that exist are small by comparison with the variation in sensitivity of the same subject in different tests.

Table 2. *Comparison of observed and expected frequencies of response*

Range of Y	Frequencies of results				
	Observed −	Observed +	Total	Expected −	Expected +
–4	8	2	10	9·72	0·28
4–5	6	0	6	3·92	2·08
5–6	5	8	13	4·26	8·74
6–	0	10	10	0·59	9·41
Total	19	20	39	18·49	20·51

Table 3. *Comparison of subjects*

Subject	Frequencies of results				
	Observed −	Observed +	Total	Expected −	Expected +
D.W.	3	6	9	4·0	5·0
V.P.W.	4	4	8	3·5	4·5
S.J.S.	12	10	22	11·0	11·0
Total	19	20	39	18·5	20·5

6. LIMITS OF ERROR

The variances of b_1 and b_2 and the covariance between them are respectively v_{11}, v_{22} and v_{12} as defined in equation (7). Hence the variance of Y, the expected probit corresponding to any pair of values x_1, x_2, is

$$V(Y) = \frac{1}{Sw} + v_{11}(x_1 - \bar{x}_1)^2 + 2v_{12}(x_1 - \bar{x}_1)(x_2 - \bar{x}_2) + v_{22}(x_2 - \bar{x}_2)^2, \tag{12}$$

where Sw is the sum of the w column in Table 1. All these variances are derived from binomial probability distributions. In the usual form of probit analysis, with a batch of subjects at each dose, the precision of the estimated relationship between dose and response is discussed as though the variation were normal, an assumption which is justifiable on account of the large numbers of individuals involved. Here, with only thirty-nine observations in all, the

assumption is less safe, but may be adopted for lack of any more trustworthy method of dealing with the data. It is unlikely to be seriously misleading, except possibly for extreme levels of the response probability, P.

Equation (12) may now be used in the assignment of fiducial limits to any one of the curves of equal probability given by equation (10). For suppose that t is the normal deviate corresponding to the significance level to be used in defining the fiducial limits, and that Y_0 is the probit of a probability P_0. Then for any values of x_1, x_2 for which

$$(Y-Y_0)^2 > t^2V(Y),$$

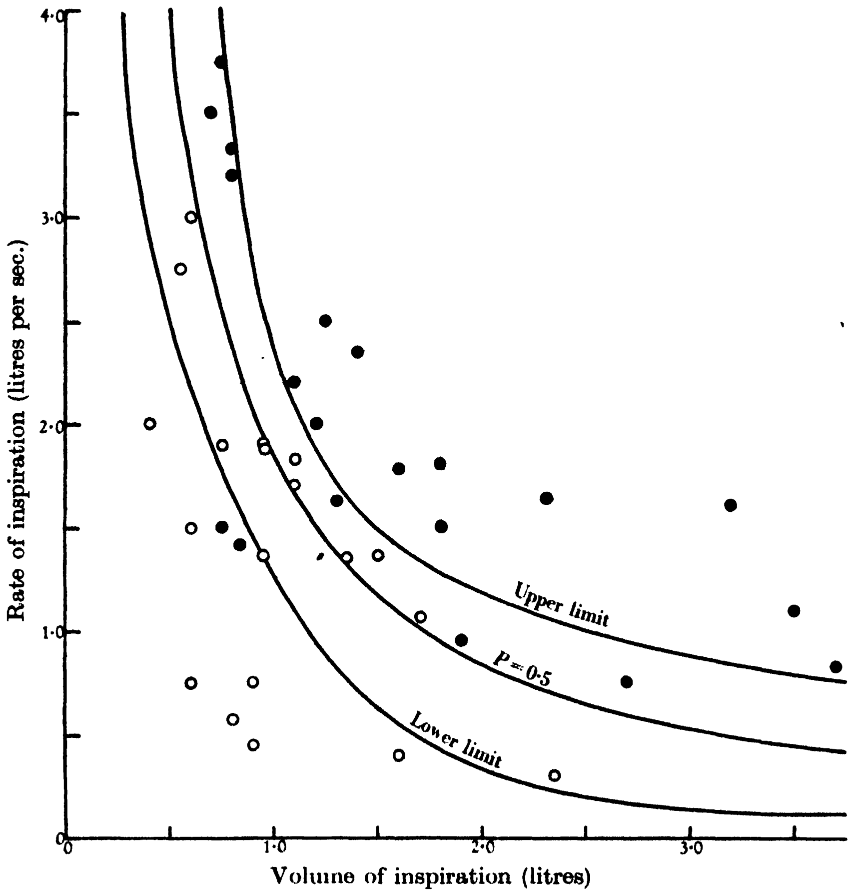


Fig. 2. Fiducial limits (5 % probability) to 0·5 frequency contour of Fig. 1.
○ no vaso-constriction; ● vaso-constriction.

where Y is determined from equation (8), the expected probit differs significantly from Y_0, and for values of x_1, x_2 which reverse the inequality the difference is not significant. Therefore the equation

$$(Y-Y_0)^2 = t^2V(Y) \tag{13}$$

gives the limiting values of (x_1, x_2) for which the null hypothesis that the true expected probit is Y_0 is not untenable in the light of the data; in other words, equation (13) defines curves in the (x_1, x_2) plane which are fiducial limits to the estimated locus of points having a constant response probability P_0. These curves are clearly hyperbolae. In Figs. 2 and 3, the 5 % fiducial limit curves ($t = 1{\cdot}960$) for $P_0 = 0{\cdot}5$ and $P_0 = 0{\cdot}9$ respectively have been plotted in

the (V, R) plane; details of the calculation need not be given here, but Fig. 2, for example, is derived from the equation

$$(14 \cdot 182 - 6 \cdot 6843 x_1 - 5 \cdot 9400 x_2)^2 = 3 \cdot 841 \left[\frac{1}{14 \cdot 29} + 4 \cdot 7261 (x_1 - 1 \cdot 0495)^2 + 6 \cdot 9878 (x_1 - 1 \cdot 0495)(x_2 - 1 \cdot 2483) + 4 \cdot 5145 (x_2 - 1 \cdot 2483)^2 \right].$$

The pairs of curves are like hyperbolae in form. That for $P_0 = 0 \cdot 5$ defines a band on either side of the estimated relationship which is quite narrow for moderate values of V and R

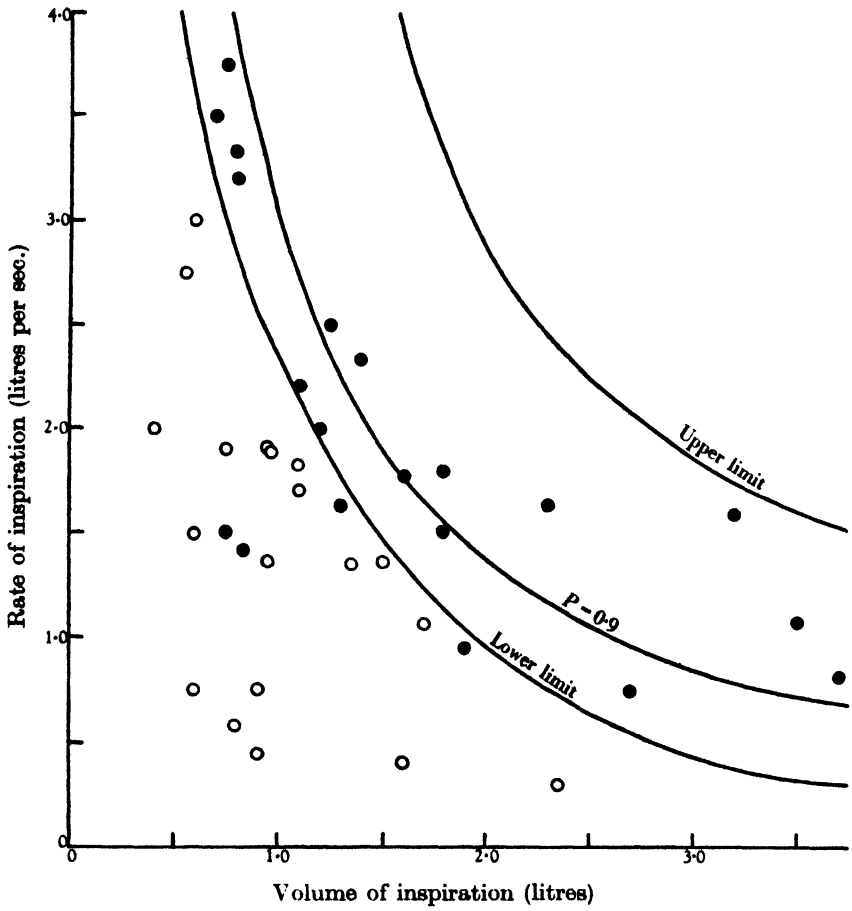


Fig. 3. Fiducial limits (5 % probability) to 0·9 frequency contour of Fig. 1.
○ no vaso-constriction; ● vaso-constriction.

though naturally it widens considerably at the extremes. That for $P_0 = 0 \cdot 9$, as might be expected from general consideration of the problem, allows much greater uncertainty on the side of high values of V and R; similarly, for $P_0 = 0 \cdot 1$, that band would be relatively wider on the side of low values of V or R.

The curves shown in Figs. 1, 2 and 3 may be regarded as plane sections, for selected values of Y, of a three-dimensional diagram relating Y to V and R. In terms of x_1, x_2 instead of V, R, this diagram is the three-dimensional analogue of the familiar diagram showing a regression line with hyperbolic curves indicating limits of error on either side; the line

generalizes to a plane, and the limits are now defined by two sheets of a hyperboloid, one above and one below the plane.

The theoretical basis of the curves illustrated in Figs. 2 and 3 is perhaps insecure, but undoubtedly they give a useful indication of the dependence of the probability of response on V and R and of the reliability of the estimation of this relationship. Much as an experimenter might wish for a more precise assessment of the effects of V and R, experience suggests that results such as those obtained here are as good as can be expected from a total of thirty-nine quantal observations.

7. The two-parameter equation

In § 3, the equation

$$VR = \text{constant} \tag{1}$$

was suggested as an expression of the curves of constant response probability, but the more complex equation (2) was adopted for use in §§ 4–6. There are no theoretical reasons for believing that equation (1) represents the true form of the relationship, and the more general form was chosen in order that the complete calculations might be illustrated. The values of b_1 and b_2 obtained, however, do not differ very greatly by comparison with the standard error of their difference; in fact

$$V(b_1 - b_2) = v_{11} - 2v_{12} + v_{22}$$
$$= 2{\cdot}253,$$

and therefore

$$b_1 - b_2 = 0{\cdot}744 \pm 1{\cdot}501.$$

In the absence of any significant difference between the regression coefficients, the common scientific procedure of preferring the simpler hypothesis (Occam's Razor) suggests that equation (4) might be replaced by

$$Y = \alpha + \beta(x_1 + x_2). \tag{14}$$

For the estimation of equation (14), the computations are similar to, but shorter than, those of § 4, since $(x_1 + x_2)$ may be replaced by a single variate, x, and a simple regression calculated; the calculations in § 4 were used to give a first set of expected probits, from which was derived the estimate

$$Y = -9{\cdot}475 + 6{\cdot}4067(x_1 + x_2). \tag{15}$$

Only two parameters have been estimated from the data, and calculation as for equation (11) gives

$$\chi^2_{[37]} = 28{\cdot}76.$$

The difference between the two χ^2 values may be taken as a further criterion of whether or not the extra parameter is needed, closely related to the test of significance of $(b_1 - b_2)$;

$$\chi^2_{[1]} = 1{\cdot}32$$

is not significant, though again the validity of the χ^2 test is in doubt.

Substitution of the probit of a specified probability in equation (15) gives the value of the constant in equation (1). For the 50 % response probability, for example, the constant is 1·82; over the range of values tested, the curves

$$VR^{0{\cdot}889} = 1{\cdot}71 \quad \text{and} \quad VR = 1{\cdot}82$$

differ only slightly. Similarly, fiducial limits to $(x_1 + x_2)$ may be calculated, for any Y_0, as upper and lower values of the product VR. No special interest attaches to these calculations; the novelties due to the individual records are exactly as for the three-parameter equation discussed in earlier sections, and otherwise the method is entirely that of ordinary probit analysis (Finney, 1947, Chapter 4). For comparison with the three-parameter equation, diagrams similar to Figs. 2 and 3 may be prepared; both the constant probability curves and

the fiducial limits are then true hyperbolae. Fig. 4 shows the results for a 50 % response probability, and is to be compared with Fig. 2. The constant probability curve in Fig. 4 differs little from that in Fig. 2, though naturally the difference increases for large values of V or R where the curves are less well determined. For moderate values of V and R, the fiducial

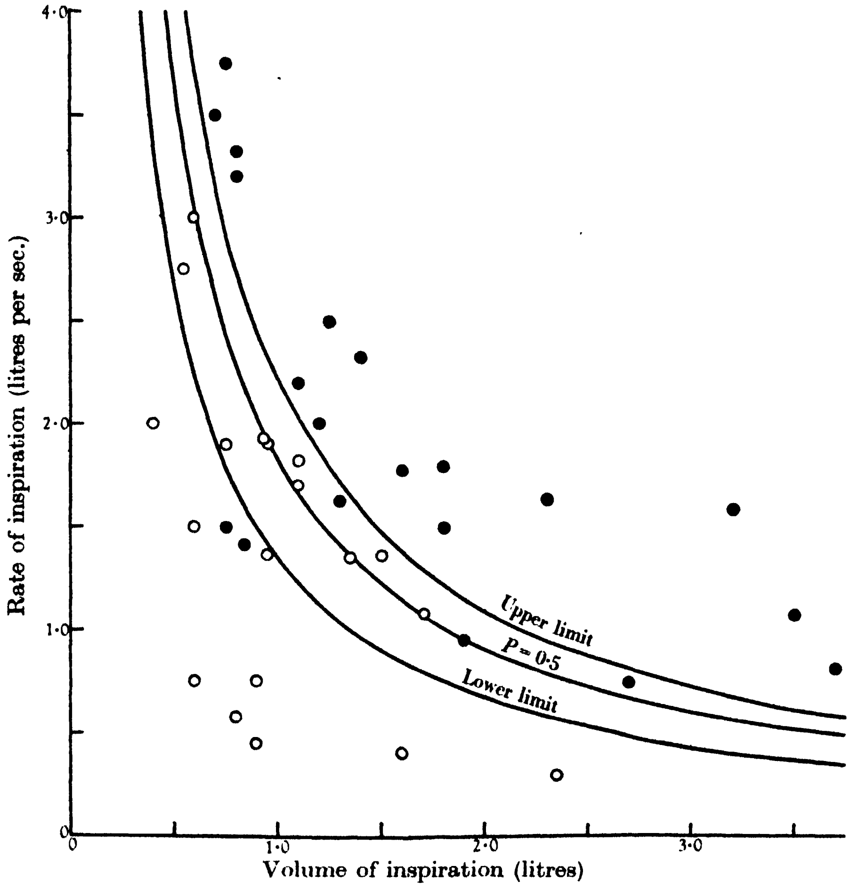


Fig. 4. Contour of dose-response surface for 0·5 frequency of response, estimated from two-parameter equation, and its 5 % fiducial limits (compare Fig. 2). ○ no vaso-constriction; ● vaso-constriction.

limit curves are practically the same as the corresponding curves in Fig. 2, but for more extreme values they lie much closer to the curve of constant probability; since the data show no significant difference between b_1 and b_2, it is to be expected that a more precisely estimated relationship between stimulus and response will be obtained if an assumption that $\beta_1 = \beta_2$ is made, so that the information on the two regression coefficients can be combined, and this shows itself by narrowing the zone of error for the constant probability curve.

8. SUMMARY

The method of probit analysis has been developed to assist the study of the relationship between the magnitude of a stimulus and the proportion of tests in which a particular quantal response to that stimulus appears. In some research problems, the stimulus cannot be controlled sufficiently to make possible the administration of a specified magnitude, though the stimulus actually received by any one subject can later be measured. It will then seldom happen that two subjects receive exactly the same 'dose', and the data for

statistical analysis will generally consist of a series of doses with, for each, a statement of whether or not a single subject showed the characteristic response.

Even for data of this type, the probit transformation can aid the estimation of the relationship between dose and the probability of response. The calculations leading to the estimate are more tedious than is usual in probit analysis, because of slow convergence from a provisional equation to the final form, but follow the usual pattern. The validity of the χ^2 test of goodness of fit (in reality a test for the normality of distribution of individual tolerances) must be doubted, however, since the disturbance due to small class numbers will be encountered in its most extreme form. Extensive grouping of results for adjacent doses will provide a test less open to objection, though this will generally be insensitive to all but the grossest deviations from normality; indeed, no valid sensitive test is to be expected with individual records unless these are very numerous.

In this paper, the calculations have been illustrated on data relating to a reflex vasoconstriction which sometimes occurs in the skin of the digits of human subjects after a single deep breath. The relationship between the occurrence of this response and two dose factors, the volume and the rate of inspiration, has been estimated for the combined records from three subjects; inclusion of two dose factors complicates the analysis, since a bi-variate regression equation must be fitted, but does not affect the underlying theory. The χ^2 test has been discussed at length, though there is no indication of non-normality or of heterogeneity of the data. The reliability with which the dependence of the probability of response on the dose factors is estimated has also been examined, and curves bounding fiducial regions, within which the true probability contours may confidently be asserted to lie, have been determined. This method of representing the limits of error is applicable to other forms of probit analysis involving two dose factors and is not restricted to individual records, though it has not previously been described.

I am indebted to Mr R. W. Gilliatt, of the Department of Physiology, both for permission to make use of his data in an illustration of the statistical methods of my paper and for assistance in describing his experimental procedure. My thanks are due also to Miss M. Callow, who prepared Figs. 1–4.

REFERENCES

Bliss, C. I. (1934*a*). The method of probits. *Science*, **79**, 38–9.

Bliss, C. I. (1934*b*). The method of probits—a correction. *Science*, **79**, 409–10.

Bliss, C. I. (1935*a*). The calculation of the dosage-mortality curve. *Ann. Appl. Biol.* **22**, 134–67.

Bliss, C. I. (1935*b*). The comparison of dosage-mortality data. *Ann. Appl. Biol.* **22**, 307–33.

Bliss, C. I. (1938). The determination of dosage-mortality curves from small numbers. *Quart. J. Pharm.* **11**, 192–216.

Bolton, B., Carmichael, E. A. & Stürup, G. (1936). Vaso-constriction following deep inspiration. *J. Physiol.* **86**, 83–94.

Finney, D. J. (1943). The statistical treatment of toxicological data relating to more than one dosage factor. *Ann. Appl. Biol.* **30**, 71–9.

Finney, D. J. (1947). *Probit Analysis: A Statistical Treatment of the Sigmoid Response Curve.* Cambridge: University Press.

Finney, D. J. (1948). The principles of biological assay. *J.R. statist. Soc. Suppl.* **9**, 46–91.

Fisher, R. A. (1946). *Statistical Methods for Research Workers*, 10th ed. Edinburgh: Oliver and Boyd.

Fisher, R. A. & Yates, F. (1947). *Statistical Tables for Biological, Agricultural and Medical Research*, 3rd ed. Edinburgh: Oliver and Boyd.

Gaddum, J. H. (1933). Reports on biological standards. III. Methods of biological assay depending on a quantal response. *Spec. Rep. Ser. Med. Res. Coun., Lond.*, no. 183.

Gilliatt, R. W. (1947). Vaso-constriction in the finger following deep inspiration. *J. Physiol.* (in the Press).

A POWER FUNCTION FOR TESTS OF RANDOMNESS IN A SEQUENCE OF ALTERNATIVES

BY F. N. DAVID

1. During recent years attention has been focused on what might be called the 'group' test for randomness in a sequence of alternatives. Thus, if E denote the happening of an event, and $\bar{E}$ its negation, the number of alternations of E and $\bar{E}$ in a sequence supposedly random has been chosen as a test criterion. This test has been put to different uses by W. L. Stevens (1939), A. Wald & J. Wolfowitz (1940) and F. N. David (1947). It seems worth while therefore to enquire what is the power of this test against a set of specifically defined alternate hypotheses. The hypothesis to be tested will be that there is randomness within the sequence, with the alternate hypothesis that if there is no randomness then there is dependence of the type found in a simple Markoff chain. The same procedure will hold good for dependence of the types found in more complex chains although in these cases the enumeration is a little troublesome.

2. If there is a sequence of dependent events

$$E_1, E_2, E_3, \ldots, E_n,$$

then it is an elementary proposition of the probability calculus that

$$P\{E_1E_2E_3\ldots E_{n-1}E_n\} = P\{E_1\}P\{E_2 \mid E_1\}P\{E_3 \mid E_1E_2\}\ldots P\{E_n \mid E_1E_2\ldots E_{n-1}\}.$$

If the events are independent, then

$$P\{E_1E_2E_3\ldots E_{n-1}E_n\} = P\{E_1\}P\{E_2\}P\{E_3\}\ldots P\{E_n\}.$$

This relation will be the basis of H_0, the hypothesis to be tested. If there is dependence as in a simple Markoff chain, then mathematically each event will be dependent on the event immediately preceding it, but will be independent of any of the other events. In this case we shall have

$$P\{E_1E_2E_3\ldots E_{n-1}E_n\} = P\{E_1\}P\{E_2 \mid E_1\}P\{E_3 \mid E_2\}\ldots P\{E_n \mid E_{n-1}\}.$$

This relation will be the basis of H_1, the hypothesis alternate to H_0.

3. For the hypothesis, H_0, let the probability that an event E will occur in a single trial be p, and let the probability of $\bar{E}$ (the negation of E), be q, where $p+q=1$. The probability of obtaining any given sequence of r_1 E's and r_2 $\bar{E}$'s will be

$$p^{r_1}q^{r_2}.$$

The number of ways in which r_1 E's and r_2 $\bar{E}$'s may be arranged to form $2t$ and $2t+1$ sets of E's and $\bar{E}$'s alternately is

$$f_{2t} = \frac{2(r_1-1)!\,(r_2-1)!}{(t-1)!\,(t-1)!\,(r_1-t)!\,(r_2-t)!} \quad \text{and} \quad f_{2t+1} = f_{2t} \times \frac{r_1+r_2-2t}{2t}.$$

Writing $k = 2t$ or $2t+1$ as desired, the probability of obtaining a sequence of r_1 E's and r_2 $\bar{E}$'s arranged in k sets is

$$P\{k \mid r_1, r_2, H_0\} = \frac{p^{r_1}q^{r_2}f_k}{\sum\limits_{\text{All } t} p^{r_1}q^{r_2}f_t} = \frac{f_k}{\sum\limits_{\text{All } t} f_t}.$$

k may take values $2, 3, \ldots, 2r_2$, if $r_1 = r_2$, and values $2, 3, \ldots, 2r_2+1$ if $r_1 > r_2$.

4. Following the orthodox procedure, in order to test the hypothesis, H_0, it is necessary to find two numbers k_1 and k_2 such that

$$P\{k < k_1 \mid H_0\} \leqslant \tfrac{1}{2}\epsilon, \quad P\{k > k_2 \mid H_0\} \leqslant \tfrac{1}{2}\epsilon,$$

and therefore
$$P\{k_1 \leqslant k \leqslant k_2\} \geqslant 1 - \epsilon,$$

where ϵ is a number arbitrarily at choice. If an observed number of sets, say k', falls outside the limits k_1 and k_2 then the hypothesis H_0 will be rejected in favour of some alternate hypothesis, H_1. Alternately if H_0 is not true, but H_1 is, then

$$1 - P\{k_1 \leqslant k \leqslant k_2 \mid H_1\}$$

will be the power of the test in the sense of the word as used by Neyman & Pearson. Whether k_1 or k_2 is chosen to judge the significance of an observed k' will depend on which departure from randomness it is most important not to overlook. If the alternate hypothesis is that there is positive dependence in the chain, i.e. that E having occurred in the sth trial it is more likely to occur in the $(s+1)$st trial, then k_1 would be chosen. Such a situation was envisaged in a proposed smooth test to supplement the χ^2 criterion (David, 1947). If, however, the alternate hypothesis is that there is negative dependence, i.e. that E having occurred in the sth trial, it is less likely to occur in the $(s+1)$st trial, then k_2 would be the appropriate criterion. If it is immaterial whether the departure from randomness is positive or negative dependence, then both k_1 and k_2 may be used.

5. We now consider the alternate hypothesis, H_1. Write E_s for the occurrence of the event E in the sth trial and $\bar{E}_s$ for its negation. Let

$$P\{E_1\} = P, \qquad P\{\bar{E}_1\} = Q, \quad P + Q = 1 \text{ and } P \geqslant Q,$$
$$P\{E_s \mid E_{s-1}\} = p_1, \quad P\{\bar{E}_s \mid E_{s-1}\} = q_1,$$
$$P\{E_s \mid \bar{E}_{s-1}\} = p_2, \quad P\{\bar{E}_s \mid \bar{E}_{s-1}\} = q_2.$$

Thus p_1 and q_2 are probabilities of no change and p_2 and q_1 probabilities of a change. If the events are independent then

$$p_1 = p_2 = P \quad \text{and} \quad q_1 = q_2 = Q.$$

6. In calculating the probability of obtaining any given sequence, what will matter will be the number of changes from E to $\bar{E}$ and back again. Let $f_t(r_1)$ be the number of ways in which r_1 E's can be arranged in t groups, i.e. let

$$f_t(r_1) = \frac{(r_1 - 1)!}{(t-1)!\,(r_1 - t)!}.$$

If there are $2t$ groups in a sequence of r_1 E's and r_2 $\bar{E}$'s, the number of ways of obtaining such a sequence will be
$$f_t(r_1)\, f_t(r_2)$$
if the sequence starts with E or with $\bar{E}$. The probability of obtaining any given sequence of r_1 E's and r_2 $\bar{E}$'s of $2t$ groups will be

$$Pp_2^{t-1}p_1^{r_1-t}q_1^t q_2^{r_2-t} \quad \text{or} \quad Qq_1^{t-1}q_2^{r_2-t}p_2^t p_1^{r_1-t}.$$

This follows from the fact that a sequence of $2t$ groups beginning with E will imply t changes from E to $\bar{E}$ and $t-1$ changes from $\bar{E}$ to E. The changes are reversed in number if the sequence starts with $\bar{E}$. For $2t+1$ groups the number of ways of obtaining the sequence will be

$$f_{t+1}(r_1)\, f_t(r_2) \quad \text{or} \quad f_t(r_1)\, f_{t+1}(r_2)$$

according as the sequence begins with E or $\bar{E}$. The respective probabilities will be

$$Pp_2^t p_1^{r_1-t-1} q_1^t q_2^{r_2-t} \quad \text{and} \quad Qq_1^t q_2^{r_2-t-1} p_2^t p_1^{r_1-t}.$$

The probability therefore of obtaining a sequence of $r_1\,E$'s and $r_2\,\bar{E}$'s in $2t$ groups will be therefore, under hypothesis H_1,

$$P\{2t \mid r_1 r_2 H_1\} = \frac{\left(\dfrac{p_2 q_1}{p_1 q_2}\right)^t f_t(r_1)\, f_t(r_2) \left(\dfrac{P}{p_2} + \dfrac{Q}{q_1}\right)}{\sum\limits_{t=1}^{r_2} \left(\dfrac{p_2 q_1}{p_1 q_2}\right)^t \left[f_t(r_1)\, f_t(r_2) \left(\dfrac{P}{p_2} + \dfrac{Q}{q_1}\right) + \dfrac{P}{p_1} f_{t+1}(r_1)\, f_t(r_2) + \dfrac{Q}{q_2} f_t(r_1)\, f_{t+1}(r_2)\right]}.$$

The probability of obtaining $r_1\,E$'s and $r_2\,\bar{E}$'s in $2t+1$ groups will be similarly

$$P\{2t+1 \mid r_1 r_2 H_1\} = \frac{\left(\dfrac{p_2 q_1}{p_1 q_2}\right)^t \left[\dfrac{P}{p_1} f_{t+1}(r_1)\, f_t(r_2) + \dfrac{Q}{q_2} f_t(r_1)\, f_{t+1}(r_2)\right]}{\sum\limits_{t=1}^{r_2} \left(\dfrac{p_2 q_1}{p_1 q_2}\right)^t \left[f_t(r_1)\, f_t(r_2) \left(\dfrac{P}{p_2} + \dfrac{Q}{q_1}\right) + \dfrac{P}{p_1} f_{t+1}(r_1)\, f_t(r_2) + \dfrac{Q}{q_2} f_t(r_1)\, f_{t+1}(r_2)\right]}.$$

7. So far no mention has been made of any possible connexion between p_1, q_1, p_2 and q_2. It is obvious in all cases we shall have

$$p_1 + q_1 = 1, \quad p_2 + q_2 = 1,$$

but the connexion between p_1 and p_2 is not immediate. We shall make the simplifying assumption which is perhaps most closely related to practical problems, and shall state that where nothing is known about the $s-1$ trials preceding the sth trial, $P\{E_s\} = P$ and $P\{\bar{E}_s\} = Q$. Under this assumption we have

$$p_2 = \frac{Pq_1}{Q}, \quad q_2 = 1 - \frac{Pq_1}{Q}.$$

This result is reached easily by noticing that

$$P\{E_s\} = P\{E_s E_{s-1}\} + P\{E_s \bar{E}_{s-1}\} = P\{E_{s-1}\}\, P\{E_s \mid E_{s-1}\} + P\{\bar{E}_{s-1}\}\, P\{E_s \mid \bar{E}_{s-1}\}$$

whence

$$P = Pp_1 + Qp_2.$$

8. The alternative hypothesis chosen to illustrate the power function formulae is that there is positive dependence in the sequence, i.e. k_1 is found so that

$$P\{k \leqslant k_1 \mid H_0\} \leqslant \epsilon \quad \text{and} \quad 1 - P\{k \geqslant k_1 \mid H_1\}$$

is calculated, when $p_1 \geqslant P$. For economy of drawing, several power curves or what are really sections of a kind of power surface, plotted to coordinates P, p_1, have been put together in the diagrams of Fig. 1. For example the bottom left-hand diagram shows for $r_1 = r_2 = 10$ sections of the conditional power surface for $P = 0{\cdot}5$, $0{\cdot}6$ and $0{\cdot}75$. When H_0 is true and $P = p_1$, we have the 5 % risk of rejecting H_0 wrongly. As $p_1 - P$ increases the chance of detecting the fact increases, but in a way dependent on P. The other three diagrams show similar sections of the surfaces with $r_1 = r_2 = 5$, with $r_1 = 14$, $r_2 = 6$ and with $r_1 = 7$, $r_2 = 3$. In practice it will not be known what the value of P is, but the curves show reasonably well how the power of the test varies as P and p_1 (and therefore p_2) vary. It is clear that the test for randomness under discussion is most powerful when the numbers of alternates are equal, i.e. when $r_1 = r_2$. The power declines sharply when r_1 increases at the expense of r_2. Another point which emerges is that the test is only moderately powerful, against the given alternate hypothesis tested, when $r_1 + r_2 = 20$, and it would appear therefore that if it was desired not to overlook a possible departure from randomness in the form of positive dependence in the chain, then the length of the sequence should consist of at least 20 units. The question of other possible tests we shall not discuss at this stage.

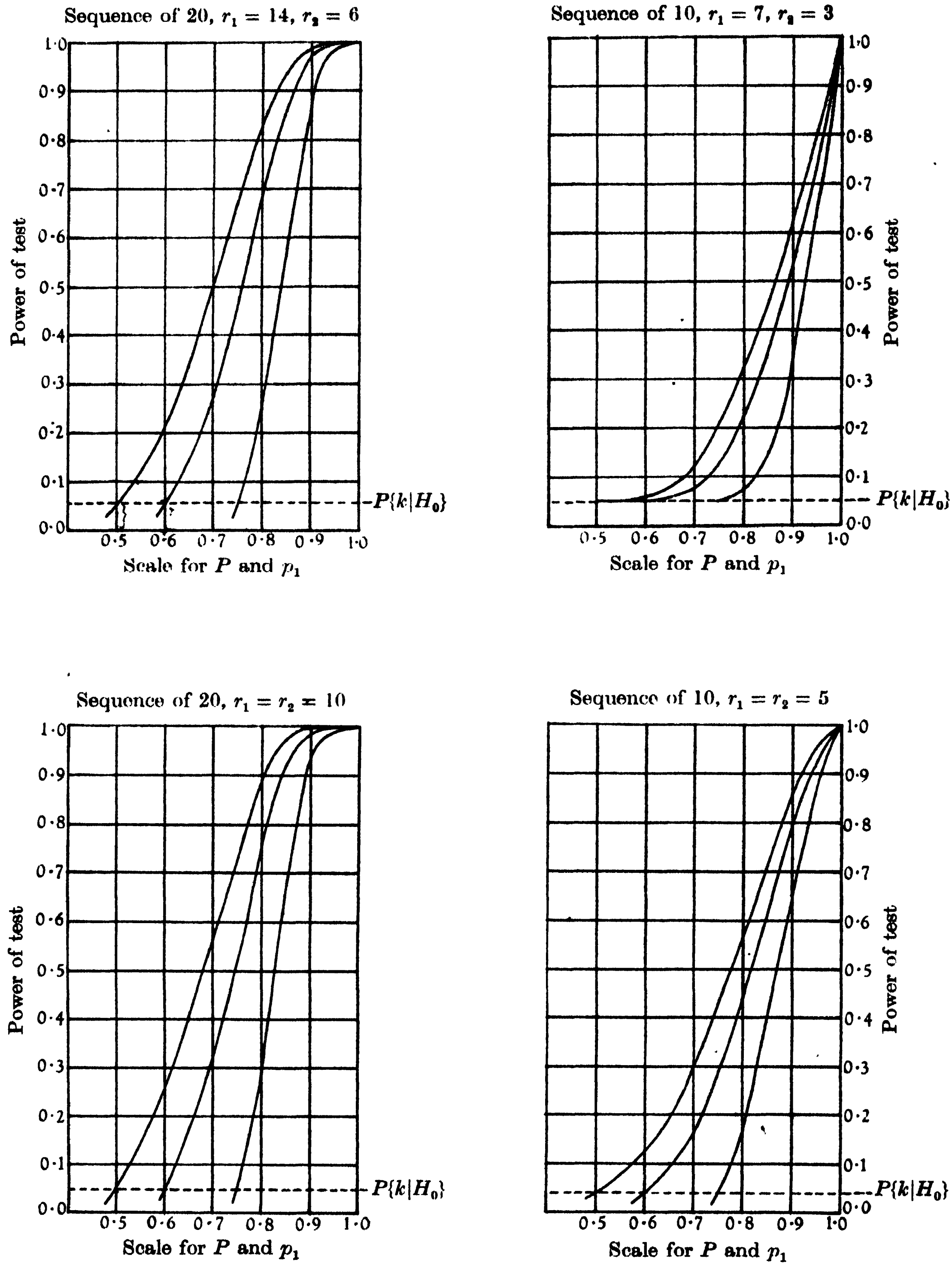


Fig. 1. Conditional power curves when the alternate hypothesis is positive dependence.

9. It will be noticed that $P\{2t \text{ or } 2t+1 \mid r_1 r_2 H_1\}$ which have been loosely termed power function formulae are not power functions in the sense originally defined by Neyman & Pearson, but they appear to involve a justifiable extension of that idea. In order to distinguish them from the usual meaning of the words power function, I shall refer to them as *conditional power functions*. The theory of the conditional power function may be stated briefly in the following way. It is assumed that all possible samples (or sequences) may be classified according to their composition. Suppose that there are k of these mutually exclusive classes, which are also the only possible, say $C_1, C_2, \ldots, C_k$. We have considered only the case where k is finite but it appears likely that the method can be extended to cover the case where k is enumerably infinite. These classes, $C_1, C_2, \ldots, C_k$ will correspond to regions forming a partition of the sample space.

Let H_0 be the hypothesis tested and w_0 be the critical region used for the rejection of this hypothesis. Given that a sample is in C_i (say), and that an alternate hypothesis H_1 is true, then the probability that H_0 will be rejected is

$$P\{E\epsilon w_0 C_i \mid E\epsilon C_i, H_1\} = \frac{P\{E\epsilon w_0 C_i \mid H_1\}}{P\{E\epsilon C_i \mid H_1\}}$$

where $w_0 C_i$ means the region common to w_0 and to C_i and, following the Neyman-Pearson notation, E is the sample point. Regarded as a function of H_1 this is the conditional power function of the test associated with w_0 in the subset C_i of samples.

The Neyman-Pearson power function, which we might call here the *overall* power function, will be

$$P\{E\epsilon w_0 \mid H_1\} = \sum_{i=1}^{k} P\{E\epsilon w_0 C_i \mid H_1\} = \sum_{i=1}^{k} P\{E\epsilon w_0 C_i \mid E\epsilon C_i, H_1\} P\{E\epsilon C_i \mid H_1\},$$

which may be looked on as a weighted average of the conditional power functions.

10. There seems to be no reason why w_0 should not be built up of portions $w_0 C_i$, these portions being chosen to maximize each term of the summation, i.e. $w_0 C_i$ chosen to maximize the conditional power function. For example, to revert to the specific case of randomness within a sequence with which we have been dealing, the different partitions of r $(= r_1 + r_2)$ are the mutually exclusive and only possible classes C_i. It is conceivable, although practically not very likely, that for each of these classes there will exist a different test which is more powerful to detect specifically defined departures from the basic hypothesis tested than any other test. The decision as to which is the most powerful test, against the same specifically defined alternatives, to use for any given class will be decided by the conditional power function. Once this has been decided the procedure for the complete test of significance may be laid down. This will be: (i) count the number of alternatives in the sequence, i.e. find r_1 and r_2, (ii) from (i) decide the appropriate test of significance to use, (iii) apply the test. The power of the test as laid down by (i), (ii) and (iii), in the usual meaning of the word, will be given by the overall power function.

It is proposed to discuss these, and other applications of the conditional power function technique, in a further publication. I have been concerned here with trying to explain what I believe to be the basic ideas, and to forestall possible criticism that I am falling into error (of the third kind) and am choosing the test falsely to suit the significance of the sample.

REFERENCES

STEVENS, W. L. (1939). *Ann. Eugen., Lond.*, **9**, 11.

WALD, A. & WOLFOWITZ, J. (1940). *Ann. Math. Statist.* **11**, 147.

DAVID, F. N. (1947). *Biometrika*, **34**, 299.

A NUMERICAL SOLUTION OF THE PROBLEM OF MOMENTS

By H. O. HARTLEY and S. H. KHAMIS

1. Introduction

Given a statistical variable x and its frequency distribution $f(x)$, then, under certain continuity conditions for $f(x)$, the moments

$$\mu_r = \int x^r f(x)\,dx \quad (r = 0, 1, 2, \ldots) \tag{1}$$

can be evaluated for any integer r. For certain distributions $f(x)$ the integrations in (1) can be carried out analytically resulting in simple formulae for the moments. In general there is no inherent difficulty in obtaining numerical values for the moments by numerical quadrature.

The inverse problem is to find the distribution $f(x)$ given the moments μ_r. This problem, commonly known as 'The Problem of Moments', has received considerable attention by mathematicians and is of interest in statistical distribution theory. There are numerous statistics for which it is difficult to obtain a formula of the random sampling distribution $f(x)$ amenable to numerical evaluation. On the other hand, in such cases it is often possible to find simple formulae for the random sampling moments (Bartlett, 1937). Sometimes such formulae are available for *all* integer r; more often than not, however, μ_r is only known for a limited number of small r (e.g. $r = 0, 1, \ldots, 6$). A simple method of 'determining' $f(x)$ from the given moments would therefore be helpful in such cases.

Examples of variables of this kind are the numerous moment statistics or k-statistics for which random sampling moments can be evaluated, notably by R. A. Fisher's (1929, 1930) combinatorial methods, whilst their exact sampling distributions are usually unknown. As related statistics we should mention here the moment ratios $\sqrt{b_1}$ and b_2 used in tests for deviation from normality (Geary, 1947, Geary & Worlledge, 1947). For these, the low-order moments are known exactly. A similar situation arises with statistics defined as likelihood ratios, as, for instance, with the criterion L_1 required for testing heterogeneity in a set of variances. Moments for this statistic were obtained by Neyman & Pearson as early as 1931, yet, although approximations to $f(L_1)$ have been obtained (Bartlett, 1937; Hartley, 1940; Nayer, 1936; Neyman & Pearson, 1931; Sukhatme, 1936; Welch, 1935, 1936), there is still considerable doubt about their accuracy in certain cases, and the exact formula obtained by Nair (1936) in the case of equal sample sizes is very complex.

These and numerous other problems of distribution point to the necessity of developing a numerical technique to deal with the following situation:

(i) A random variable x ranging between a and b (where a may be $-\infty$ and b may be $+\infty$) has a distribution function $f(x)$ known to have a continuous derivative of order n.

(ii) The moments
$$\mu_r = \int_a^b x^r f(x)\,dx \quad (r = 0, 1, \ldots, R), \tag{2}$$

are known numerically to any decimal accuracy desired but for a limited number of positive integers r, viz. $r = 0, 1, \ldots, R$. With the knowledge about $f(x)$ limited to the above conditions, is it possible to obtain numerical values for the probability integral $P(x) = \int_a^x f(x)\,dx$

depending on the moments only, and is it possible to make a statement on the accuracy of these values in terms of the derivatives of the function $f(x)$?

Problems of this kind have hitherto been treated principally in two ways:

(a) When $R = 2, 3$ or 4 nothing better can be expected than a 'good fit', which is often achieved by fitting the appropriate Pearson-type curve.

(b) With R in the neighbourhood of 5–8, expansions of the Gram Charlier, Laguerre or Jacobi type have been used, either as cumulant or as moment expansions. Such theorems as are available for statements on the convergence and asymptotic behaviour of these expansions usually require too many moments to be known. Often the expansions are only asymptotic, and unless the distribution is close to the generating curve (Normal for Gram Charlier, Γ for Laguerre), the results are often disappointing (see, for example, Kendall, 1945, Chapter 6).

2. OUTLINE OF PRESENT METHOD

The method to be developed here is a direct application of finite-difference calculus and therefore provides both numerical answers to the problem, as well as gauges of their accuracy in form of remainder terms. The method is, in fact, closely linked with interpolation technique. When using any of the well-known interpolation formulae no mathematically rigorous statement on the accuracy of the interpolates can be made unless the magnitude of the remainder term can be estimated, and for this some knowledge about (say) the nth derivative of the function is required. Yet, in using such formulae the convergence of the difference table inspires confidence that 'the results of the interpolation can be accepted as a working hypothesis' (Milne Thomson, 1933, p. 62). Similarly, with the present method we shall give a numerical procedure of obtaining values of the probability integral. Certain checks of internal consistency will be described which inspire confidence that the answers are correct, but no rigorous statement on the accuracy can, of course, be made if this is to be based on a finite number of moments alone. The *exact* remainder terms which we derive will entail the high-order derivatives of $f(x)$, and it is hoped, in a second communication, to derive some *general* statements concerning their order of magnitude.

In order to simplify the argument we assume in this section that the range of x is finite (a and b finite).

The aim is to determine the probability integral of x, $P(x) = \int_a^x f(x)\,dx$ in tabular form, i.e. we wish to determine numerical values of

$$P_i = P(x_i) = \int_a^{x_i} f(x)\,dx \tag{3}$$

for discrete values of x_i. For convenience the group intervals $x_{i+1} - x_i$ will generally be chosen equidistant (group interval $= h$), and the number of intervals will be $R+1$, i.e. equal to the number of given moments (including $\mu_0 = 1$). Hence

$$x_i = a + ih, \quad h = (b-a)/(R+1). \tag{4}$$

The first differences in the table derived from equation (3) are the quantities

$$f_i = P_i - P_{i-1} = \int^{x_i} f(x)\,dx, \tag{5}$$

and are the familiar 'frequencies' f_i in a grouped frequency distribution with equidistant intervals (see Fig. 1). The link between these frequencies and the exact moments μ_r is then

established by the well-known formulae for Sheppard's correction. Using Kendall's (1938, 1945) derivation and remainder term, but extending his notation, we have

$$\sum_{i=1}^{R+1} f_i \xi_i^r = \mu_r + C(r,h) + S(r,h), \tag{6}$$

where the centre points ξ_i are given by

$$\xi_i = a + (i - \tfrac{1}{2})h \quad (i = 1, \ldots, R+1), \tag{7}$$

$C(r,h)$ denotes Sheppard's corrective term, viz.

$$C(r,h) = \sum_{j=1}^{[\frac{1}{2}r]} \left(\frac{h}{2}\right)^{2j} \binom{r}{2j} \frac{1}{2j+1} \mu_{r-2j}, \tag{8}$$

and $S(r,h)$ the remainder term.

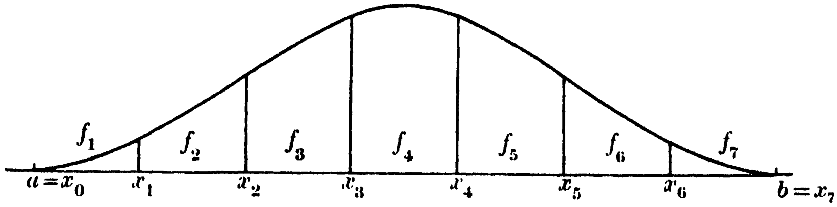


Fig. 1

The aim, now, is to use equations (6) to determine the unknown f_i from the given μ_r. To this end the remainder $S(r,h)$ must be examined: Most distribution functions have what is commonly known as high contact at the terminals of the variate range. This means that $f(x)$, as well as all its derivatives up to order, say, m, vanish at both ends of the range, i.e.

$$f^{(i)}(a) = f^{(i)}(b) = 0 \quad (i = 0, \ldots, m). \tag{9}$$

If for such functions we define $f(x) = 0$ outside the range $a \leqslant x \leqslant b$, it will have continuous derivatives of up to order m for $-\infty < x < +\infty$. It can then be shown that the remainder term is of the form (see, for example, Kendall, 1945, p. 69)

$$S_m(r,h) = -\frac{(R+1)h^m}{m!} B_m k^{(m)}(r,h,\theta_r) \quad (m \text{ even}), \tag{10}$$

$$S_m(r,h) = \frac{2(R+1)h^m}{m!} B^{(1)}_{m+1}(\tfrac{1}{2})\, k^{(m)}(r,h,\theta_r) \quad (m \text{ odd}), \tag{11}$$

$$a \leqslant \theta_r \leqslant b,$$

where the B_j are the Bernoulli numbers, the $B^{(1)}_j$ are the Bernoulli polynomials of first order, the integrand function $k(r,h,x)$ is defined by

$$k(r,h,x) = x^r \int_{-\frac{1}{2}h}^{+\frac{1}{2}h} f(x+\xi)\, d\xi, \tag{12}$$

and its derivatives with regard to x are denoted by $k^{(i)}$. In the subsequent sections we shall assume (9) to hold (contact of order m), but will discuss the case when (9) is not satisfied in § 10.

The remainder term $S_m(r,h)$ will usually be small (see, for example, Kendall, 1945, p. 72). We shall therefore, in what follows, ignore $S_m(r,h)$ but will discuss the error thereby committed in § 5.

If, then, in (6) we omit $S(r,h)$ we obtain a system of $R+1$ linear equations for the $R+1$ unknowns f_i

$$\sum_{i=1}^{R+1} f_i \xi_i^r = \mu_r + C(r,h) = \bar{\mu}_r. \tag{13}$$

The matrix of this system of equations (v_R say) is of the form $|\xi_i^r|$ and has a classical determinant $\|\xi_i^r\|$, sometimes referred to as Vandermonde's determinant and well known to be $\neq 0$. The system can therefore be inverted once and for all and, for any *particular* case, the unknown f_i can then be determined by substituting the right-hand sides of (13), i.e. $\bar{\mu}_r$ in the inverse matrix v_R^{-1}. Denoting the elements of this inverse matrix by u_{ir} we have the system of equations

$$f_i = \sum_{r=0}^{R} u_{ir}\bar{\mu}_r. \tag{14}$$

Progressive addition of the f_i yields the P_j from $P_j = \sum_{i=1}^{j} f_i$* and therefore a table of $P(x)$ at interval h. Finally, intermediate values of $P(x)$ can be obtained by standard interpolation. Alternatively, as described in § 7, we may obtain directly a table of $P(x)$ at interval $\frac{1}{2}h$.

3. The standard form of the numerical inversion

The rank of the original matrix v_R is obviously equal to $R+1$, i.e. the number of moments given, whilst its elements are the powers of the centre points ξ_i^r. It is desirable therefore that, for any given R, scale and location of the variable x be transformed into a standard form X, so that only *one* matrix V_R and therefore only *one* matrix V_R^{-1} need be calculated for each R. It is most convenient to standardize as follows:

$$X = (x - \tfrac{1}{2}(a+b))\frac{R+1}{b-a} \quad (R \text{ even}), \tag{15}$$

$$X = (x - \tfrac{1}{2}(a+b))\frac{R+1}{b-a} + \frac{1}{2} \quad (R \text{ odd}). \tag{16}$$

It will be seen, therefore, that the range of X is $R+1$ and the group interval

$$H = X_{i+1} - X_i = 1.$$

From the given moments of x those of X (M_r say) about $X = 0$ can, of course, be calculated by the usual binomial formulae, and in what follows we assume that values of M_r are given numerically. Further, in analogy to (13), we have

$$\bar{M}_r = M_r + C(r, 1). \tag{17}$$

From (15) and (16) we obtain for the new centre points

$$\left.\begin{aligned} \Xi_i &= -\tfrac{1}{2}R, \ldots, 0, \ldots, +\tfrac{1}{2}R \quad \text{for even } R, \\ \Xi_i &= -\frac{R-1}{2}, \ldots, 0, \ldots, \frac{R+1}{2} \quad \text{for odd } R, \end{aligned}\right\} \tag{18}$$

and the matrix V_R becomes $|(i-1-\frac{1}{2}R)^r|$ or $|(i-\frac{1}{2}R-\frac{1}{2})^r|$. Thus, if the first six moments are given, we obtain for V_6:

$$V_6 \equiv \begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ 9 & 4 & 1 & 0 & 1 & 4 & 9 \\ -27 & -8 & -1 & 0 & 1 & 8 & 27 \\ 81 & 16 & 1 & 0 & 1 & 16 & 81 \\ -243 & -32 & -1 & 0 & 1 & 32 & 243 \\ 729 & 64 & 1 & 0 & 1 & 64 & 729 \end{vmatrix}. \tag{19}$$

* It is, of course, possible to construct a matrix yielding the P_j directly from the $\bar{\mu}_r$, but we are here satisfied with determining the f_i first, as they are of independent interest.

In practice the important range of R will be from 5 to about 8. The inverse matrix V_6^{-1} is given below, and it is hoped to give V_7^{-1}, V_8^{-1} and V_5^{-1} in a subsequent paper. The inverse matrix V_6^{-1}, the elements of which are denoted by U_{ir}, can be written in the form

$$c_i f_i = \sum_{r=0}^{R} U'_{ir} \bar{M}_r, \tag{20}$$

where $U'_{ir} = c_i U_{ir}$, i.e. the c_i are suitable common denominators of the $U_{i0}, \ldots, U_{iR}$, and the U'_{ir} are given in the body of the schedule below:

		$\bar{M}_0 = 1$	$\bar{M}_1$	$\bar{M}_2$	$\bar{M}_3$	$\bar{M}_4$	$\bar{M}_5$	$\bar{M}_6$ = multiplier of column
i	$c_i f_i$	$r = 0$	1	2	3	4	5	6
1	$720f_1$	0	−12	4	15	−5	−3	1
2	$120f_2$	0	18	−9	−20	10	2	−1
3	$48f_3$	0	−36	36	13	−13	−1	1
4	$36f_4$	36	0	−49	0	14	0	−1
5	$48f_5$	0	36	36	−13	−13	1	1
6	$120f_6$	0	−18	−9	20	10	−2	−1
7	$720f_7$	0	12	4	−15	−5	3	1

(21)

In order to use the above system of equations it would be necessary to compute the $\bar{M}_r$ from the given M_r, using formula (17). It is obviously more convenient to evaluate, once and for all, a matrix U''_{ir} giving the f_i directly in terms of the given M_r. This matrix is given below for $R = 6$:

i	f_i	$M_0 = 1$	M_1	M_2	M_3	M_4	M_5	M_6
1	f_1	0·000 379	−0·011 719	0·002 344	0·017 361	−0·005 208	−0·004 167	0·001 389
2	f_2	−0·005 227	0·109 375	−0·034 896	−0·152 778	0·072 917	0·016 667	−0·008 333
3	f_3	0·059 161	−0·683 594	0·618 490	0·253 472	−0·244 792	−0·020 833	0·020 833
4	f_4	0·891 373	0	−1·171 875	0	0·354 167	0	−0·027 778
5	f_5	0·059 161	0·683 594	0·618 490	−0·253 472	−0·244 792	0·020 833	0·020 833
6	f_6	−0·005 227	−0·109 375	−0·034 896	0·152 778	0·072 917	−0·016 667	−0·008 333
7	f_7	0·000 379	0·011 719	0·002 344	−0·017 361	−0·005 208	0·004 167	0·001 389

(22)

Working rule: Each f_i is obtained by forming the sum of seven products using the seven coefficients in the ith line and applying them to $M_0, \ldots, M_6$, e.g. $f_1 = 0{\cdot}000\,379M_0 - 0{\cdot}011\,719M_1 + \ldots + 0{\cdot}001\,389M_6$.

4. Calculation of the incomplete B-function $I_x(8, 6)$ from its first six moments

As an example for the above method we consider the Beta Distribution for $p = 8$ and $q = 6$, viz.

$$f(x) = [B(8, 6)]^{-1} x^7 (1-x)^5.$$

Using the moments for this distribution about $x = 0$, $\mu_r = B(x+r, 6)/B(8, 6)$ $(r = 0, \ldots, 6)$ and transforming to the standard scale $X = 7x - 3{\cdot}5$, we obtain for the moments of X about $X = 0$: $M_1 = 0{\cdot}5$, $M_2 = 1{\cdot}05$, $M_3 = 1{\cdot}225$, $M_4 = 2{\cdot}77426$, $M_5 = 4{\cdot}41360$ and $M_6 = 10{\cdot}56942$. Substituting these in the matrix (22) we obtain values of f_i whose progressive sums are shown in Table 1 (calculated $I_x(8, 6)$). These may be compared with the 'exact' values obtained (by interpolation) from the *Tables of the Incomplete B-function* (1934). The worst discrepancy is about 2 in the fourth decimal. Higher accuracy can, of course, be obtained if the number of moments $(R+1)$ and therefore the number of f_i increases (see, for example, § 8, where the normal curve is obtained to 5-decimal accuracy).

A rather gratifying feature of the comparison is the higher *decimal* accuracy in the tails of the distribution. This is a consequence of the sensitivity of the higher moments to changes in the tail frequencies. Note also that the elements in the top and bottom lines of the inverse matrix (22) are much smaller than those in the other lines, so that any error in the right-hand sides of (13) has a smaller effect on the terminal f_t.

Table 1. *Comparison of 'calculated' and 'exact' values of* $I_x(8, 6)$

X	x	Exact I_x	Calculated I_x	Difference 10^{-5}
$-2\cdot5$	1/7	0·000 11	0·000 09	2
$-1\cdot5$	2/7	0·013 41	0·013 54	-13
$-0\cdot5$	3/7	0·140 17	0·139 95	22
0·5	4/7	0·489 63	0·489 81	-18
1·5	5/7	0·862 61	0·862 70	-9
2·5	6/7	0·994 11	0·993 95	16
3·5	7/7	1·000 00	1·000 00	0

It might be argued that a further error will arise when determining intermediate values of I_x by interpolation in the 'calculated' table. This difficulty could, however, be overcome by shifting the grid of group intervals and using a standard X-scale with group end-points corresponding to the odd multiples of 1/14 in x, thereby obtaining I_x at points half-way between the arguments of Table 1. Such a method has actually been used in § 7.

5. THE REMAINDER TERM

A formal representation of the remainder term is immediately obtained by reverting to the exact equations (6). If we are concerned with distribution functions having contact of order m at the terminals, the error contributions to the f_t are obtained by substituting the $R+1$ remainder terms $S_m(r, h)$ ((10), (11)) in the inverse matrix v^{-1}. It is convenient to use the standard variate X-scale, $H = 1$ and the V^{-1} matrix when it will be found that

$$\text{error } f_t = \sum_{r=0}^{R} U_{tr} S_m(r, 1), \tag{23}$$

where $S_m(r, 1)$ is given by (10) or (11) putting $h = 1$ and remembering that the integrand function k must be taken in terms of the standard variate X, viz.

$$k(r, 1, X) = X^r \int_{-\frac{1}{2}}^{+\frac{1}{2}} f\left(\frac{b-a}{R+1}(X+\xi)\right)\frac{b-a}{R+1}\,d\xi. \tag{24}$$

Since the arguments θ_r of $k^{(m)}(r, 1, X)$ are unknown it will as a rule be necessary to substitute their respective maxima in (23), at the same time taking $|U_{tr}|$ in place of U_{tr}.

Although with (23) we have given a formal solution of the error term involved, in a manner similar to the remainder terms of interpolation formulae, it will in practice be difficult to estimate the magnitude of the error from this formula. It is hoped, therefore, to go into this aspect more fully in a second paper.

6. INFINITE VARIATE RANGE AND ARTIFICIAL TRUNCATION

When the range of the variate is infinite, i.e. when $a = -\infty$ and/or $b = +\infty$, it is, of course, possible to transform the variate x by, say, $y = y(x)$ such that the range of y is finite. However, in general, we shall not be able to assume that the moments of y are known or that they can

be derived from those of x. It is therefore necessary to adapt our method to deal with an infinite variate range. We shall treat here the case $b = +\infty$, the case $a = -\infty$ being identical and the case $a = -\infty$ *and* $b = +\infty$ being analogous.

For an infinite variate range, the condition of high contact is now replaced by

$$\lim_{x\to\infty} f^{(i)}(x) = 0 \quad (i = 0, 1, \ldots, m), \tag{25}$$

which results in remainder terms analogous to (10) and (11)*. Similarly, in equations (6) which correspond to Kendall's (1945) equations (3·40), the summation now extends from $i = 1$ to $i = \infty$, there being an infinity of frequencies $f_i = \int_{x_{i-1}}^{x_i} f(x)\,dx$. Now since the μ_r exist we know that

$$\int_a^\infty x^r f(x)\,dx \tag{26}$$

is convergent. Accordingly

$$\lim_{b\to\infty} \sum_{i=R+2}^{\infty} (i-\tfrac{1}{2})^r h^r \int_{(i-1)h}^{ih} f(x)\,dx = 0, \tag{27}$$

if $h = (b-a)/(R+1)$. If, therefore, we denote the above sums by $\epsilon(r, b)$ respectively we have, from (6),

$$\sum_{i=1}^{R+1} f_i \xi_i^r + \epsilon(r, b) = \mu_r + C(r, h) + S(r, h). \tag{28}$$

Applying now the previous method we introduce an additional error in the calculation of f_i, but this error is smaller than

$$+\max |\epsilon(r, b)| \,\Sigma\, |u_{ir}|.$$

The precise determination of the $\epsilon(r, b)$ for any given b would, of course, require a knowledge of the nature of the convergence in (26), i.e. some external knowledge about the distribution $f(x)$ which we are seeking to determine numerically. Unfortunately, such knowledge will in general not be available.

However, if b is chosen sufficiently large, the f_i determined for *different* values of b should all yield, by the method of §§ 2 and 3, approximations to the *same* probability integral $P(x)$ to within the errors of the respective remainder terms $S(r, h)$ and to within the errors introduced by (27). In practice, therefore, one would make an intelligent guess at the likely range of b and then test for internal consistency by comparing the probability integral tables obtained by varying b over this range. This method, which is illustrated in § 7 gives an idea of the accuracy to which the integral has been determined, but no rigorous statement on accuracy can be made without appealing to some *a priori* knowledge about $f(x)$. It is hoped to deal with this aspect more fully in the next paper.

7. The calculation of the χ-distribution for 10 degrees of freedom

As an illustration of the preceding section, we will now calculate the χ-distribution for 10 degrees of freedom. This distribution has high contact at either terminal and, although it is known to start at $x = \chi = 0$, we shall treat it as a distribution of double infinite range, i.e. we shall not make direct use of the information that $f(x) = 0$ for $x \leqslant 0$, and choose a *truncated* range $a \leqslant x \leqslant b$.

We have a mean of $\mu_1 = 3{\cdot}0843\,2776$, and the moments about the mean are given by†

$$\mu_2' = 0{\cdot}486\,9223, \quad \mu_3' = 0{\cdot}080\,6720, \quad \mu_4' = 0{\cdot}713\,2999, \quad \mu_5' = 0{\cdot}386\,6784, \quad \mu_6' = 1{\cdot}810\,4865.$$

* A formula for $S(r, h)$ when the range is infinite will be given in the second paper.

† These follow from the formulae for the moments about the origin which are ratios of Γ-functions (see, for example, Kendall, 1945, p. 55). Note that we have used μ and μ' for moments about the origin and the mean, respectively.

The standard deviation is $\sqrt{\mu_2'} = 0{\cdot}7$, and with seven group intervals available to cover the essential range we should choose h of the order of the standard deviation.* Our first attempt is, therefore, (a) $h = 0{\cdot}8$.

(a) If we make the mean of x the centre point of the innermost interval we have for the truncated range $a = \mu_1 - 3{\cdot}5 \times 0{\cdot}8 = \mu_1 - 2{\cdot}8$ and $b = \mu_1 + 2{\cdot}8$. For the standard variate X, the origin $X = 0$ will coincide with the mean of x and its range will be $-3{\cdot}5 \leqslant X \leqslant +3{\cdot}5$. Calculation of the moments (M_r) of X and substitution in the matrix (22) yields the following answers for the frequencies f_i:

$$f_1 = 0{\cdot}000\,5,\quad f_2 = 0{\cdot}033\,25,\quad f_3 = 0{\cdot}262\,66,\quad f_4 = 0{\cdot}424\,71,$$

$$f_5 = 0{\cdot}231\,96,\quad f_6 = 0{\cdot}042\,06,\quad f_7 = 0{\cdot}004\,87.$$

The calculated frequency (f_7) for the interval $\mu_1 + 2{\cdot}0 \leqslant x \leqslant \mu_1 + 2{\cdot}8$ is about 0·005, and its contribution to μ_6' about $0{\cdot}005\,(2{\cdot}4)^6 \sim 1$. Since this is an appreciable proportion of μ_6' it is unlikely that the frequencies beyond $b = \mu_1 + 2{\cdot}8$ when substituted in (27) can be neglected, i.e. b and h are too small.†

(b) Choosing therefore a larger h, we try $h = 1$. If we still keep the mean in the centre of the truncated range we have $a = \mu_1 - 3{\cdot}5 = -0{\cdot}42$ and $b = \mu_1 + 3{\cdot}5 = 6{\cdot}58$ (we know, of course, that $f(x) = 0$ for $x = 0$ so that our f_1 will really be the frequency for the interval $0 \leqslant x \leqslant 0{\cdot}085$). This time the standard variate is $X = x - \mu_1$ so that $M_r = \mu_r'$, and the above values can be substituted directly in the matrix (22) yielding the comparison of calculated χ-integral and 'exact' χ-integral as shown in Table 2.

Table 2. *Comparison of calculated and exact values of the χ-integral*

$=\chi-\mu_1$	$P(x)$ exact	$P(x)$ calculated	Difference 10^{-5}
−2·5	0·000 06	0·000 11	−5
−1·5	0·009 29	0·008 93	36
−0·5	0·244 66	0·244 75	−9
+0·5	0·767 67	0·767 85	−18
+1·5	0·979 02	0·978 88	14
+2·5	0·999 45	0·999 47	−2
+3·5	1·000 00	1·000 00	0

The maximum error is about 0·0004 and, again, the terminal f_i have a higher decimal accuracy. In practice, of course, the exact distribution would not be available for comparison. This time the terminal value f_7 is about 0·0005 and represents the frequency for the interval $\mu_1 + 2{\cdot}5 \leqslant x \leqslant \mu_1 + 3{\cdot}5$. Its contribution to μ_6' is about 0·4, thereby confirming that the previous grid of group intervals was too fine. To obtain further confirmation on the tail of the distribution, we determine a third set of f_i by shifting the grid of group intervals by 0·5 to the right, retaining the interval $h = 1$. This will make $a = \mu_1 - 3$ and $b = \mu_1 + 4$, i.e. $0{\cdot}08 \leqslant x \leqslant 7{\cdot}08$. For our standard variate X the origin will now coincide with $\mu_1 + 0{\cdot}5$.

* An unsuitable choice of h would, later, fail to satisfy the checks of internal consistency.

† Comparison with the exact χ-distribution shows that the maximum error in the above f_i is nevertheless not more than 0·005.

The values of the M_r are as follows:

$$M_1 = -0{\cdot}5, \quad M_2 = 0{\cdot}736\,9223, \quad M_3 = -0{\cdot}774\,7114,$$
$$M_4 = +1{\cdot}344\,8394, \quad M_5 = -1{\cdot}834\,794, \quad M_6 = 3{\cdot}595\,7606.$$

Substituting these in the matrix (22) we obtain the following values of f_i:

$$f_1 = 0{\cdot}000\,15, \quad f_2 = 0{\cdot}070\,12, \quad f_3 = 0{\cdot}444\,30, \quad f_4 = 0{\cdot}404\,21,$$
$$f_5 = 0{\cdot}076\,99, \quad f_6 = 0{\cdot}004\,19, \quad f_7 = 0{\cdot}000\,04.$$

The comparison of the progressive sums of the above f_i with the exact χ-integral is of similar accuracy to that in Table 2. The terminal frequency for $\mu_1 + 3 \leqslant x \leqslant \mu_1 + 4$ is 0·0005 with a contribution of about 0·03 to μ_6', indicating that we have now reached a satisfactory choice of b.

As a final check on the internal consistency we compare the answers obtained with the two last choices of group intervals by merging the tables of $P(x)$ to obtain one table at interval 0·5. This is set out in Table 3. The differences provide a fair check on the internal consistency to about 3-decimal accuracy of the two separate tables. If a more reliable check is desired, three or even four separate tables may be computed, all at the same group interval h and merged in the above manner to form a single table at interval $\frac{1}{3}h$ or $\frac{1}{4}h$. This procedure has the added advantage that interpolation difficulties at the wide interval of h are being avoided.

Table 3. $x = \chi$ *for 10 degrees of freedom. Calculated table of $P(x)$ obtained from two separate grids of group intervals* ($h = 1$)

$x-\mu_1$	$P(x)$			
−2·5	0·0001			
		1		
−2·0	0·0002		86	
		87		442
−1·5	0·0089		528	
		615		601
−1·0	0·0704		1129	
		1744		−174
−0·5	0·2448		955	
		2699		−1122
0	0·5147		−167	
		2532		−855
0·5	0·7679		−1022	
		1510		112
1·0	0·9189		−910	
		600		480
1·5	0·9789		−430	
		170		296
2·0	0·9959		−134	
		36		103
2·5	0·9995		−31	
		5		
3·0	1·0000			

8. The special case of symmetrical distributions; the normal integral

By placing the origin of the standard variate X at the mean of a symmetrical distribution we obviously have $f_1 = f_{R+1}$, $f_2 = f_R$, etc., i.e. the number of unknowns is halved. On the other hand, the odd moments contribute the meaningless equations

$$\Sigma f_i(\Xi_i^r + \Xi_{R+2-i}^r) = \Sigma f_i \times 0 = 0.$$

With the number of unknowns and equations halved and with even moments only retained, it is necessary to work out a new matrix ($\hat{V}_R$ say) based on even-order moments only. In practice the important values of R are $R = 4, 6, 8$ and 10, and we are giving below the inverse matrix $\hat{V}_8^{-1}$ (for $R = 8$) having rank 5 (as there are five equations corresponding to μ_0, μ_2, μ_4, μ_6 and μ_8):

i	f_i	$M_0 = 1$	M_2	M_4	M_6	M_8
1	f_1	0·000 3441	−0·001 7857	0·001 2153	−0·000 2315	0·000 0124
2	f_2	−0·003 9874	0·020 8333	−0·013 7153	0·002 3148	−0·000 0868
3	f_3	0·022 4151	−0·119 0476	0·071 1806	−0·008 1019	0·000 2480
4	f_4	−0·088 4281	0·500 0000	−0·143 4028	0·012 7315	−0·000 3472
5	f_5	0·569 6563	−0·400 0000	0·084 7222	−0·006 7130	0·000 1736

(29)

Working rule: Each f_i is obtained by forming the sums of five products using the five coefficients in the ith line and applying them to $M_0, \ldots, M_8$; e.g. $f_1 = 0{\cdot}000\,3441 M_0 - \ldots + 0{\cdot}000\,0124 M_8$.

As an example we compute the normal integral from its first five even moments, μ_0 to μ_8, choosing $h = 1$ and the standard variate X as normal deviate. Substituting, therefore, in the matrix (29) $M_0 = 1$, $M_2 = 1$, $M_4 = 3$, $M_6 = 15$ and $M_8 = 105$, we obtain the five f_i which in Table 4 have been progressively added to form the 'calculated normal integral' to be compared with the 'exact' one. The accuracy is remarkable, the maximum error being 15 in the 6th decimal.

Table 4. *Comparison of calculated normal integral with exact normal integral*

$x = X$	Exact $P(x)$	Calculated $P(x)$	Difference $\times 10^{-6}$
−4	0·000 032	0·000 034	−2
−3	0·001 350	0·001 342	8
−2	0·022 750	0·022 765	−15
−1	0·158 655	0·158 643	12
0	0·500 000	0·500 000	0

With symmetrical distributions we cannot, of course, shift the grid of group intervals, as otherwise we would lose the symmetry relation between the f_i. If, therefore, intermediate values of $P(x)$ are required in order to ease subsequent interpolation, we can achieve this only by altering h. Merging the answers obtained from (say) three different h grids all centred at $x = 0$ (e.g. $h = 0{\cdot}9$, $1{\cdot}0$ and $1{\cdot}1$), we would *not* obtain a table of $P(x)$ at an *equidistant* interval. In the internal check we would, therefore, use divided differences.

9. Divergent or poorly convergent moments; the t-distribution for 10 degrees of freedom

Some variates with infinite range have distribution functions with low contact at $x = \infty$, i.e. the convergence in

$$\lim_{x \to \infty} f(x) = 0 \tag{30}$$

is slow, indeed, in some cases the moment μ_r is divergent for, say, $r \geqslant R'$.

As an example we have investigated the t-distribution for 10 degrees of freedom. Here we have $f(x) = c(1 + t^2/10)^{-5\cdot5}$ and hence $R' = 10$. In this case, therefore, R' is known *a priori*. If no such mathematical information is available, warning of low contact is given by the rapid

growth of the moments as $r \to R$, provided R is near to R'.* For our example for the t-distribution we find

$$\mu_2 = 1\cdot25, \quad \mu_4 = 6\cdot25, \quad \mu_6 = 78\cdot125, \quad \mu_8 = 2734\cdot375.$$

The difficulty with such distributions is that artificial truncation is not justified if the high-order, poorly convergent moments are to be used in equations (6). The remedy in such cases is the square variate transformation $y^2 = x$. Sometimes it may be necessary to use a higher power $y^k = x$. Obviously, if we were to take an equidistant interval for y, the group integral for x will *grow* with the square law, thereby absorbing the slowly convergent tail end of $f(x)$.

Now, obviously, the moments of y are simply related to those of x; we have

$$\int_0^\infty x^r f(x)\,dx = 2\int_0^\infty y^{2r} y f(y^2)\,dy, \tag{31}$$

or introducing the new distribution function $g(y) = 2yf(y^2)$, we have

$$\int_0^\infty x^r f(x)\,dx = \int_0^\infty y^{2r} g(y)\,dy. \tag{32}$$

Applying now the previous method to $g(y)$ it is further necessary to avoid using the poorly convergent high-order moments. In the case of the t-distribution, instead of taking $r = 0, 2, 4, 6$ and 8, we take the absolute moments† for $r = 0, 1, 2, 3$ and 4, which, according to (32), correspond to the even moment of $g(y)$. If only even moments about the origin are used in the determination of the f_i, the matrix (29) gives the appropriate inversion. Using $h = 0\cdot6$ for the y-group interval we substitute in (29):

$M_0 = 1, \quad M_2 = 2\cdot401\,906, \quad M_4 = 9\cdot645\,062, \quad M_6 = 52\cdot952\,032$ and $M_8 = 372\cdot108\,863.$

We thereby obtain five values of f_i ($i = 1, \dots, 5$) of the form

$$f_i = \int_{(5-i)h}^{(6-i)h} g(y)\,dy = \int_{(5-i)^2h^2}^{(6-i)^2h^2} f(x)\,dx. \tag{33}$$

The progressive sums of these are compared with the corresponding values of the exact t-integral in Table 5. Although the accuracy is lower than in the previous example it is satisfactory and very much better than we could have obtained without applying the transformation $y^2 = x$.

Table 5. *Comparison of calculated and exact values of the t-integral*

t	$P(t)$ exact	$P(t)$ calculated	Difference $\times 10^{-4}$
5·76	0·0001	0·0001	0
3·24	0·0044	0·0042	2
1·44	0·0902	0·0905	− 3
0·36	0·3613	0·3648	−35
0·00	0·5000	0·5000	0

* If R is much smaller than R', the present difficulty will not arise at all.

† We shall show in a second paper that, if the absolute moments of a distribution are not known, they can be obtained by interpolation between the values of $\log \mu_r$ for $r = 2, 4, 6, 8$, etc.; in fact, we shall give a general discussion of the interpolability of the logarithmic moment function for positive x.

10. Lack of high contact at the start of the variate $x = a$

We confine ourselves here to the most important case of lack of high contact at one terminal, say the start of the distribution, and assume, therefore, that there *is* high contact at *one end of the range*.

Without loss of generality we assume that $a = 0$, i.e. $x \geqslant 0$, and introduce the new variate $y^k = x$, $k \geqslant 2$. Whence we have

$$\int_0^b x^r f(x)\,dx = \int_0^{b^{1/k}} y^{kr} g(y)\,dy, \tag{34}$$

where $g(y) = ky^{k-1}f(y^k)$. Obviously $g(y)$ has, at least, contact of order $k-1$ at the start $y = 0$; further, if $f(x)$ has contact of order m at $x = b$, i.e. if $f(x) = O(x^{-m})$ at $x = b$, then $g(y) = O(y^{-(m-1)k-1})$. Hence there is high-order contact, of order $k-1$ and $(m-1)k+1$, respectively, at both ends of the range. The previous method is therefore applicable to $g(y)$ provided we can obtain its moments from those of $f(x)$. It is obvious from (34) that in order to obtain the ordinary moment of $g(y)$ we require to know the 'fractional' moments of $f(x)$, i.e. those corresponding to $r = j/k$ ($j = 0, 1, \ldots$). If the moments of $f(x)$ are only known for integer r the fractional moments will have to be obtained by interpolation of the logarithmic moment function $\log \mu_r$ which will be more fully discussed in the next paper.

REFERENCES

Bartlett, M. S. (1937). *Proc. Roy. Soc.* A, **160**, 268.
Fisher, R. A. (1929). *Proc. Lond. Math. Soc.* (2), **30**, 199.
Fisher, R. A. (1930). *Proc. Roy. Soc.* A, **130**, 16.
Geary, R. C. (1947). *Biometrika*, **34**, 68.
Geary, R. C. & Worlledge, J. P. G. (1947). *Biometrika*, **34**, 98.
Hartley, H. O. (1940). *Biometrika*, **31**, 249.
Kendall, M. G. (1938). *J.R. statist. Soc.* **101**, 592.
Kendall, M. G. (1945). *The Advanced Theory of Statistics*, **1**. London: C. Griffin and Co.
Milne Thomson, L. M. (1933). *The Calculus of Finite Differences*. London: Macmillan.
Nair, U. S. (1936). *Biometrika*, **30**, 274.
Nayer, P. P. N. (1936). *Statist. Res. Mem.* **1**, 38.
Neyman, J. & Pearson, E. S. (1931). *Bull. int. Acad. Cracovie*, A, p. 460.
Sukhatme, P. V. (1936). *Statist. Res. Mem.* **1**, 94.
Tables of the Incomplete Beta-Function, edited by Karl Pearson (1934). London: Biometrika Office.
Welch, B. L. (1935). *Biometrika*, **27**, 145.
Welch, B. L. (1936). *Statist. Res. Mem.* **1**, 1.

APPROXIMATION TO PERCENTAGE POINTS OF THE z-DISTRIBUTION

By A. H. CARTER, *King's College, Cambridge*

Tables have been published of the values of z for various percentage levels (20, 5, 1 and 0·1 %) for a range of given n_1, n_2 (Fisher & Yates, 1943, Table V). When n_1 or n_2 is outside the range of the tables, recourse must be had to approximate formulae (unless, of course, interpolation is sufficiently accurate) which will combine accuracy with facility of computation. One such formula, due to Fisher, with a modification suggested by Cochran (1940), is given at the foot of the above-mentioned tables. The purpose of this paper is to derive an alternative formula, no more difficult to compute, which will be shown to give consistently closer approximations to the true value of z for all except small n_1 or n_2.

Wishart (1947) has derived formulae for the exact cumulants of z, and also the well-known approximations to them when n_1, n_2 are large. The exact cumulants as far as κ_5 can be readily obtained arithmetically from tables of the Polygamma functions. Knowing the cumulants of the distribution, we may make use of the Cornish-Fisher normalization function method, based on Edgeworth's form of the Gram-Charlier type A series (Cornish & Fisher, 1937), to approximate to the percentage points. The method consists in writing z as an expansion in powers of a corresponding normal variate, ξ, the coefficients being functions of the cumulants of z, and assumes that κ_r is of order n^{1-r}, which is true for the z-distribution (Wishart, 1947, p. 172).

If z and ξ are expressed in standard measure (i.e. mean zero, standard deviation unity) we then derive

$$\frac{z-\mu_1'}{\sigma} = z' \sim \xi + \frac{\kappa_3}{\sigma^3}\frac{\xi^2-1}{6} + \frac{\kappa_4}{\sigma^4}\frac{\xi^3-3\xi}{24} - \frac{\kappa_3^2}{\sigma^6}\frac{2\xi^3-5\xi}{36},$$

correct to order n^{-1} for z'. This gives

Formula (a):
$$z \sim \mu_1' + \sigma\xi + \frac{\kappa_3}{\sigma^2}\frac{\xi^2-1}{6} + \frac{\kappa_4}{\sigma^3}\frac{\xi^3-3\xi}{24} - \frac{\kappa_3^2}{\sigma^5}\frac{2\xi^3-5\xi}{36}, \tag{1}$$

correct to order $n^{-\frac{3}{2}}$ (since $\sigma = O(n^{-\frac{1}{2}})$), where $\mu_1'(=k_1)$, $\sigma^2(=\kappa_2)$, κ_3, κ_4 are cumulants of the z-distribution. The ξ-coefficients may be readily computed: e.g. for the 5 % level, substitute $\xi = 1·64485$. Table 2 gives the values, for the 20, 5, 1 and 0·1 % levels, of the coefficients required in applying the formula. The quantities μ_1', σ, κ_3, κ_4 depend of course on n_1 and n_2, and may be evaluated in any particular case, whence substitution in (1) gives the appropriate value of z. Since $|z_{1-P}(n_1, n_2)| = |z_P(n_2, n_1)|$, where z_P is the value of z corresponding to probability P, to find the percentage points for the 'negative tail', i.e. 80, 95, 99 and 99·9 %, we may simply interchange n_1 and n_2. This has the effect of changing the sign of the *odd* cumulants, so that in (1) we write $-\mu_1'$ and $-\kappa_3$ for μ_1' and κ_3.

Formula (a), being an approximation to order $n^{-\frac{3}{2}}$, may be expected to give reliable results when n_1 and n_2 are both large. For the 1 % point, for example, we find $z(6, 12) = 0·7843$ (true value 0·7864), whereas $z(24, 60) = 0·3744$ (true value 0·3746). Some further results for (6, 12), (6, 60), and (24, 60) are shown in Table 1 (a).

In practice, some labour is involved in applying formula (a), even if polygamma tables are available. The Fisher-Cochran formula, derived by the normalization function method, is a simple working approximation, valid for large n_1, n_2, in which the exact cumulants are replaced by their approximations in terms of inverse powers of n_1 and n_2.

Table 1. *Comparison of approximations to the percentage points of z*

Formula (*b*): Existing formula (Fisher-Cochran).
Formula (*c*): New formula.

Percentage level	$n_1, n_2 \rightarrow$	6, 12	6, 60	24, 60	20, 36	20, 100	36, 60	24, 24	36, 36
20	Formula (*b*)	0·2687	0·1901	0·1335	0·1577	0·1287	0·1212	0·1741	0·1415
	Formula (*c*)	0·2733	0·2020	0·1340	0·1580	0·1298	0·1213	0·1740	0·1415
	True z	0·2706	0·1965	0·1338	0·1579	0·1294	0·1213	0·1740	0·1415
5	Formula (*b*)	0·5507	0·3990	0·2650	0·3128	0·2573	0·2390	0·3426	0·2778
	Formula (*c*)	0·5501	0·4100	0·2654	0·3129	0·2586	0·2391	0·3426	0·2778
	True z	0·5487	0·4064	0·2654	0·3129	0·2583	0·2391	0·3425	0·2778
1	Formula (*b*)	0·7992	0·5646	0·3746	0·4441	0·3619	0·3385	0·4894	0·3955
	Formula (*c*)	0·7886	0·5698	0·3746	0·4435	0·3629	0·3384	0·4893	0·3955
	True z	0·7864	0·5687	0·3746	0·4435	0·3630	0·3384	0·4890	0·3954
0·1	Formula (*b*)	1·1074	0·7474	0·4963	0·5928	0·4755	0·4503	0·6602	0·5307
	Formula (*c*)	1·0693	0·7372	0·4954	0·5906	0·4756	0·4498	0·6595	0·5304
	True z	1·0628	0·7377	0·4955	0·5905	0·4760	0·4498	0·6589	0·5302

Percentage level	$n_1, n_2 \rightarrow$	12, 6	60, 6	60, 24	36, 20	100, 20	60, 36
20	Formula (*b*)	0·3509	0·3346	0·1566	0·1783	0·1656	0·1314
	Formula (*c*)	0·3506	0·3408	0·1569	0·1785	0·1665	0·1315
	True z	0·3510	0·3388	0·1568	0·1784	0·1661	0·1315
5	Formula (*b*)	0·6884	0·6435	0·3047	0·3483	0·3208	0·2566
	Formula (*c*)	0·7001	0·6706	0·3060	0·3493	0·3236	0·2570
	True z	0·6931	0·6596	0·3055	0·3488	0·3227	0·2568
1	Formula (*b*)	1·0120	0·9444	0·4368	0·4995	0·4615	0·3661
	Formula (*c*)	1·0370	0·9956	0·4391	0·5016	0·4662	0·3667
	True z	1·0218	0·9770	0·4385	0·5009	0·4666	0·3666
0·1	Formula (*b*)	1·4352	1·3340	0·5930	0·6789	0·6303	0·4932
	Formula (*c*)	1·4681	1·4155	0·5965	0·6820	0·6375	0·4942
	True z	1·4449	1·3929	0·5962	0·6814	0·6371	0·4940

Table 1 (*a*). *Some values of z from formula* (*a*) (*exact cumulant formula*)

(For corresponding true values, see Table 1)

% \ n_1, n_2	6, 12	6, 60	24, 60	12, 6	60, 6	60, 24
20	0·2699	0·1998	0·1338	0·3499	0·3335	0·1567
5	0·5457	0·4022	0·2652	0·6958	0·6627	0·3057
1	0·7843	0·5640	0·3744	1·0295	0·9854	0·4388
0·1	1·0684	0·7433	0·4956	1·4592	1·4026	0·5966

The cumulant function of z is

$$K(z) = \tfrac{1}{2}it\log\frac{n_2}{n_1} + \log\Gamma\left(\frac{n_1+it}{2}\right) + \log\Gamma\left(\frac{n_2-it}{2}\right) - \log\Gamma\left(\frac{n_1}{2}\right) - \log\Gamma\left(\frac{n_2}{2}\right),$$

and the cumulants are obtainable by differentiating this successively with respect to (it), at each stage putting $t = 0$.

Since
$$\left[\frac{d^r}{d(it)^r}\log\Gamma\left(\frac{n\pm it}{2}\right)\right]_{t=0} = (\pm 1)^r\frac{d^r}{dn^r}\log\Gamma\left(\frac{n}{2}\right),$$

and $\log\Gamma(\tfrac{1}{2}n)$ can be expanded by Stirling's theorem in inverse powers of n, the cumulants may also readily be expressed in inverse powers of n_1, n_2; and, when n_1, n_2 are reasonably large, the first few terms only in the expansions will give sufficiently close approximations to them. In fact, writing
$$\frac{1}{n_1}+\frac{1}{n_2} = s, \quad \frac{1}{n_1}-\frac{1}{n_2} = d,$$

it has been shown that
$$\left.\begin{aligned}
\kappa_1 &= \mu_1' = -\tfrac{1}{2}d - 6sd + O(n^{-4}),\\
\kappa_2 &= \sigma^2 = \tfrac{1}{2}s + \tfrac{1}{4}(s^2+d^2) + O(n^{-3})\\
\kappa_3 &= -\tfrac{1}{2}sd + O(n^{-3}),\\
\kappa_4 &= \tfrac{1}{4}s(s^2+3d^2) + O(n^{-4}),\\
\kappa_r &= O(n^{1-r}) \quad (r>1).
\end{aligned}\right\} \tag{A}$$

Formula (a) will now have an extra term, since we take as our 'working variance' of z not its exact value, but its approximation to order n^{-1} from (A), i.e. $\tfrac{1}{2}s$. In the notation of Kendall (1945)
$$l_2 = \kappa_2/\tfrac{1}{2}s - 1 \sim \tfrac{1}{2}(s+d^2/s).$$

We then obtain
$$z-\mu_1' \sim \sqrt{\left(\frac{s}{2}\right)}\,\xi - \frac{d}{6}(\xi^2-1) + \sqrt{\frac{s}{2}}\left\{\frac{s}{24}(\xi^3+3\xi) + \frac{d^2}{72s}(\xi^3+11\xi)\right\} \tag{2}$$
$$= \frac{\xi}{\sqrt{2/s}}\left\{1+\frac{1}{2/s}\,\frac{\xi^2+3}{12}\right\} - \frac{d}{6}(\xi^2-1) + \frac{d^2}{144}\sqrt{\frac{2}{s}}\,(\xi^3+11\xi)$$

or
$$z-\mu_1' \sim \frac{\xi}{\sqrt{(h-\lambda)}} - \frac{d}{6}(\xi^2-1), \tag{3}$$

where
$$h = \frac{2}{s}, \quad \lambda = \frac{\xi^2+3}{6},$$

provided $\tfrac{1}{144}d^2\sqrt{(2/s)}\,(\xi^3+11\xi)$ may be neglected (which will be the case for small d). Inserting the approximation to μ_1' from (A), i.e. $-\tfrac{1}{2}d$,
$$z \sim \frac{\xi}{\sqrt{(h-\lambda)}} - \frac{d}{6}(\xi^2+2), \tag{3a}$$

the Fisher-Cochran formula, which has, in fact, been found to give a fairly close approximation to the true z for n_1, n_2 both reasonably large. It may be noted that if n_1, n_2 are not very large, an improvement will be effected by including the second term in the estimate of the mean (κ_1), i.e. from (A) by adding $-\tfrac{1}{6}sd$.

For $(n_1, n_2) = (6, 12)$ this correction is $-0{\cdot}00347$, and for $(24, 60)$, the correction is $-0{\cdot}00024$. Inserting this improved approximation to μ_1' in (3) we have

Formula (b):
$$z \sim \frac{\xi}{\sqrt{(h-\lambda)}} - \frac{d}{6}(\xi^2+2+s). \tag{3b}$$

As pointed out by Wishart (1947, p. 179), an approximation to the value of any $\kappa_r (r > 1)$ obtained by considering its leading term only, will be improved by writing $1/(n_1 - 1)$ and $1/(n_2 - 1)$ in place of $1/n_1$ and $1/n_2$. For, by Stirling's expansion of a factorial,

$$\log \Gamma\left(\frac{n}{2}\right) = \frac{n-1}{2}\log\frac{n}{2} - \frac{n}{2} + \tfrac{1}{2}\log 2\pi + \frac{1}{6n} + O(n^{-3}),$$

$$\frac{d}{dn}\log \Gamma\left(\frac{n}{2}\right) = \tfrac{1}{2}\log n - \frac{1}{2n} - \frac{1}{6n^2} + O(n^{-4}),$$

$$\frac{d^2}{dn^2}\log \Gamma\left(\frac{n}{2}\right) = \frac{1}{2n} + \frac{1}{2n^2} + \frac{1}{3n^3} + O(n^{-5})$$

$$= \frac{1}{2(n-1)} + O(n^{-3}),$$

$$\frac{d^3}{dn^3}\log \Gamma\left(\frac{n}{2}\right) = -\frac{1}{2(n-1)^2} + O(n^{-4}),$$

and so on.

Table 2. *ξ-coefficients required in applying formula* (a)

	20 %	5 %	1 %	0·1 %
ξ	0·84162	1·64485	2·32635	3·09023
$\frac{1}{6}(\xi^2-1)$	−0·04861	0·28426	0·73532	1·42492
$\frac{1}{24}(\xi^3-3\xi)$	−0·08036	−0·02018	0·23379	0·84332
$\frac{1}{36}(2\xi^3-5\xi)$	−0·08377	0·01878	0·37634	1·21026

Thus writing $$s' = \frac{1}{n_1-1} + \frac{1}{n_2-1}, \quad d' = \frac{1}{n_1-1} - \frac{1}{n_2-1},$$

we might expect a better approximation to z to be obtained, corresponding to that of Fisher and Cochran, if we use s' and d' instead of s and d.

Corresponding to equations (A) we have

$$\left.\begin{aligned} \kappa_2 &= \tfrac{1}{2}s' - \tfrac{1}{24}s'(s'^2 + 3d'^2) + O(n^{-4}),\\ \kappa_3 &= -\tfrac{1}{2}s'd' + O(n^{-4}),\\ \kappa_4 &= \tfrac{1}{4}s'(s'^2 + 3d'^2) + O(n^{-5}),\\ \kappa_r &= O(n^{1-r}) \quad (r > 1). \end{aligned}\right\} \qquad \text{(B)}$$

For the mean, however, $$\mu_1' = \kappa_1 = -\tfrac{1}{2}d - \tfrac{1}{6}sd + O(n^{-4}), \qquad (4)$$

$$= -\tfrac{1}{2}d' + \tfrac{1}{3}s'd' + O(n^{-3}), \qquad (4a)$$

If n^{-3} is *not* negligible (relative to the degree of accuracy desired) μ_1' should therefore be left in the form $-\frac{1}{2}d - \frac{1}{6}sd$.

Proceeding as before, we obtain

$$z - \mu_1' \sim \sqrt{\left(\frac{s'}{2}\right)}\,\xi - \frac{d'}{6}(\xi^2 - 1) + s'\sqrt{\left(\frac{s'}{2}\right)}\frac{\xi^3 - 3\xi}{24} + d'^2\sqrt{\left(\frac{2}{s'}\right)}\frac{\xi^3 - 7\xi}{144}, \qquad (5)$$

whence $$z - \mu_1' \sim \frac{\xi}{\sqrt{(h' - \lambda')}} - \frac{d'}{6}(\xi^2 - 1), \qquad (6)$$

where $$h' = 2/s', \quad \lambda' = \tfrac{1}{6}(\xi^2 - 3).$$

Since this is based in the first place on more accurate approximations to the cumulants κ_3 and κ_4, and since the term omitted from (5) in deriving (6) (i.e. $\frac{1}{144}d'^2\sqrt{(2/s')}(\xi^3-7\xi)$, is evidently numerically less than the corresponding term omitted in obtaining the Fisher-Cochran formula (i.e. $\frac{1}{144}d^2\sqrt{(2/s)}(\xi^3+11\xi)$, formula (6) might be expected to give an improved approximation to z. In fact, however, it does not, and the reason is not far to seek.

Consider the expansion of $\xi/\sqrt{(h-\lambda)}$ in both cases:

$$\frac{\xi}{\sqrt{(h-\lambda)}} = \frac{\xi}{\sqrt{h}}\left(1+\frac{\lambda}{2h}+\frac{3\lambda^2}{4h^2}+\dots\right),$$

where the terms are decreasing in magnitude (since $1/h = \frac{1}{2}s = O(n^{-1})$). Hence the error in neglecting all terms after the second will be approximately of the order of the third term. Now in the Fisher-Cochran approximation this term, $+\dfrac{3\lambda^2\xi}{4h^2\sqrt{h}}$, has the same sign as the omitted term $\frac{1}{144}d^2\sqrt{(2/s)}(\xi^3+11\xi)$ (both being of the same sign as ξ), so that the extra terms included will tend to compensate for the term omitted. In obtaining (6) from (5), on the other hand, the term omitted, $\frac{1}{144}d'^2\sqrt{(2/s')}(\xi^3-7\xi)$, will be of *opposite* sign to ξ when $|\xi|<\sqrt{7}$, corresponding to a probability of about 0·004: so that for most percentage levels encountered in practice, the error in (5) is increased in (6).

A better formula is obtained from (5) as

$$z-\mu_1' \sim \frac{\xi\sqrt{(h'+\lambda')}}{h'}-\frac{d'}{6}(\xi^2-1), \tag{7}$$

where h' and λ' are as in (6).

Expanding $\xi\sqrt{(h'+\lambda')}/h'$, it is found that the third term is now of *opposite* sign to ξ, and hence the extra terms contained in the expansion will tend to compensate for the term omitted. Since s' and d' require to be calculated in applying this formula, it is desirable to write μ_1' in the form (4a) (provided we can neglect quantities of order n^{-3}). This gives

Formula (*c*):
$$z \sim \frac{\xi\sqrt{(h'+\lambda')}}{h'}-\frac{d'}{6}(\xi^2+2-2s'). \tag{7a}$$

Collecting the results, we have the three approximate formulae:

Formula (*a*) (exact cumulant method):

$$z \sim \kappa_1+\sigma\xi+\frac{\kappa_3}{\sigma^2}\frac{\xi^2-1}{6}+\frac{\kappa_4}{\sigma^3}\frac{\xi^3-3\xi}{24}-\frac{\kappa_3^2}{\sigma^5}\frac{2\xi^3-5\xi}{36}.$$

Formula (*b*) (Fisher-Cochran formula):

$$z \sim \frac{\xi}{\sqrt{(h-\lambda)}}-\frac{d}{6}(\xi^2+2+s),$$

where
$$s = \frac{1}{n_1}+\frac{1}{n_2},\quad d = \frac{1}{n_1}-\frac{1}{n_2},\quad h = \frac{2}{s},\quad \lambda = \frac{\xi^2+3}{6}.$$

Formula (*c*) (new formula):
$$z \sim \frac{\xi\sqrt{(h'+\lambda')}}{h'}-\frac{d'}{6}(\xi^2+2-2s'),$$

where
$$s' = \frac{1}{n_1-1}+\frac{1}{n_2-1},\quad d' = \frac{1}{n_1-1}-\frac{1}{n_2-1},\quad h' = \frac{2}{s'},\quad \lambda' = \frac{\xi^2-3}{6}.$$

When n_1, n_2 are large (and not too different), $\frac{1}{3}sd$ is negligible and formula (b) becomes the formula more generally quoted

$$z \sim \frac{\xi}{\sqrt{(h-\lambda)}} - \left(\frac{1}{n_1} - \frac{1}{n_2}\right)\frac{\xi^2+2}{6}$$

which may be written
$$z \sim \frac{\xi}{\sqrt{(h-\lambda)}} - \left(\frac{1}{n_1} - \frac{1}{n_2}\right)(\lambda - \tfrac{1}{6}).$$

Similarly, for sufficiently large n_1, n_2, $\frac{1}{3}s'd'$ may be neglected and formula (c) becomes

$$z \sim \frac{\xi\sqrt{(h'+\lambda')}}{h'} - \left(\frac{1}{n_1-1} - \frac{1}{n_2-1}\right)\frac{\xi^2+2}{6},$$

or
$$z \sim \frac{\xi\sqrt{(h'+\lambda')}}{h'} - \left(\frac{1}{n_1-1} - \frac{1}{n_2-1}\right)(\lambda' + \tfrac{5}{6}).$$

It is to be noted, however, that since $\frac{1}{3}s'd'$ is approximately twice $\frac{1}{3}sd$, more care must be exercised in deciding to neglect it. For example, when (n_1, n_2) is $(20, 100)$, $\frac{1}{3}s'd' = 0{\cdot}0009$, and for $(24, 60)$, its value is $0{\cdot}0005$.

For purposes of comparison, values of z have been computed from formulae (b) and (c), for the four common percentage levels, over a fairly wide range of n_1, n_2. They are shown in Table 1, together with the corresponding true values of z. The latter were obtained where possible from the tables of Fisher and Yates: elsewhere by inverse interpolation in *Tables of the Incomplete Beta-Function* followed by a logarithmic transformation. Such values are in error by not more than 0·0001. It will be seen that neither formula yields very accurate results when n_1 or n_2 is as small as 6, though even here the new formula is rather better with the single exception of $n_1 = 12$, $n_2 = 6$. In actual practice, however, we are concerned with large values of n_1, n_2, beyond the range of the published tables. Considering only those cases where n_1 and n_2 are both greater than 20, it is seen that formula (c) gives a consistently closer approximation than does formula (b) for both the positive and the negative tails, and for all the percentage levels investigated, though its relative gain in accuracy is greatest at the 1 and 0·1 % levels. It may be noted, in fact, that in no case considered having n_1 and n_2 greater than 20, is the error more than 9 in the fourth decimal place, i.e. it appears that for all except small n_1, n_2, this formula will give an approximation to z correct to within 0·001.

In conclusion, therefore, it is recommended that formula (c) be adopted for general use, since it is no more difficult to compute, and is more accurate, than the existing formula. Dropping the dashes we have the formula

$$z \sim \frac{\xi\sqrt{(h+\lambda)}}{h} - \left(\frac{1}{n_1-1} - \frac{1}{n_2-1}\right)\left(\lambda + \frac{5}{6} - \frac{s}{3}\right),$$

where
$$s = \frac{1}{n_1-1} + \frac{1}{n_2-1}, \quad h = \frac{2}{s}, \quad \lambda = \frac{\xi^2-3}{6}$$

or, if $\frac{1}{3}\left\{\frac{1}{(n_1-1)^2} - \frac{1}{(n_2-1)^2}\right\}$ may be neglected,

$$z \sim \frac{\xi\sqrt{(h+\lambda)}}{h} - \left(\frac{1}{n_1-1} - \frac{1}{n_2-1}\right)(\lambda + \tfrac{5}{6}).$$

The values of ξ and λ for the four percentage levels are:

	20 %	5 %	1 %	0·1 %
ξ	0·8416	1·6449	2·3263	3·0902
λ	−0·3819	0·0491	0·4020	1·0916

My thanks are due to Dr J. Wishart, whose suggestion was the basis of this paper.

REFERENCES

Cochran, W. G. (1940). Note on an approximative formula for significance levels of z. *Ann. Math. Statist.* **11**, 93.

Cornish, E. A. & Fisher, R. A. (1937). Moments and cumulants in the specification of distributions. *Rev. Inst. Int. Statist.* **4**, 307.

Fisher, R. A. & Yates, F. (1943). *Statistical Tables*, 2nd ed. Oliver and Boyd.

Kendall, M. G. (1945). *Advanced Theory of Statistics*, **1**, 156, 2nd ed. C. Griffin and Co.

Wishart, J. (1947). The cumulants of the z and of the logarithmic χ^2 and t distributions. *Biometrika*, **34**, 170.

MISCELLANEA

Note on the cumulants of Fisher's z-distribution

By LEO A. AROIAN, *Hunter College*

In a recent article Dr J. Wishart (1947) stated: 'Explicit expressions for the exact cumulants of Fisher's z-distribution do not appear ever to have been published.' Fisher's z-distribution and the related Snedecor's F-distribution formed a part of my doctor's thesis and rather full results concerning the cumulants of the z-distribution and other properties of the distribution were published in the *Annals of Mathematical Statistics* (Aroian, 1941) some time ago.* I should like to take this opportunity of adding certain comments on the Gram-Charlier Type A approximation to the z-distribution and the type III approximation to the F-distribution.

To obtain the cumulants of the z-distribution I expanded the moment generating function $M_z(\theta)$ in powers of θ and found $\lambda_{k:z}$, the kth semi-variant (or cumulant) of z as the coefficient of $\theta^k/k!$. The exact results correspond with Wishart's formulae (9) to (15), although given in a different form, and need not be repeated here. In addition, asymptotic formulae for $\lambda_{k:z}$, n_1 and n_2 large, were derived by means of the Euler-Maclaurin sum formula. Furthermore, another type of formula could have been given for n_1 small but n_2 large, merely by expanding that part of $\lambda_{k:z}$ in which n_2 occurs by the Euler-Maclaurin sum formula. The special cases for the logarithmic χ^2, the logarithmic t, and the logarithmic normal probability functions follow by substituting the proper limiting values of n_1 and n_2.

In my previous paper I was overcautious concerning the type A approximation to the z-distribution. Actually the method is fairly accurate although tedious. Taking

$$F(t) = \phi(t) + A_3\phi^{\text{iii}}(t) + A_4\phi^{\text{iv}}(t), \quad \int_{t_0}^{\infty} F(t)\,dt = \eta,$$

we have

$$\int_{t_0}^{\infty} F(t)\,dt = \int_{t_0}^{\infty}\phi(t)\,dt + \phi(t)\{-A_3(t_0^2-1) + A_4(t_0^3-3t_0)\},$$

where η is usually 0·10, 0·05, 0·025, 0·01, etc. As an example take $n_1 = 24$, $n_2 = 60$; then

$$\lambda_{1:z} = -0{\cdot}0127429, \quad \sigma_z = 0{\cdot}173779, \quad \lambda_{3:z} = -0{\cdot}0007998, \quad \lambda_{4:z} = 0{\cdot}0000867,$$

$$A_3 = \frac{-\lambda_{3:z}}{3!\,\sigma_z^3} = 0{\cdot}025345, \quad A_4 = \frac{\lambda_{4:z}}{4!\,\sigma_z^4} = 0{\cdot}00396.$$

t_0 for $\eta = 0{\cdot}05$ is 1·60094, $z_{0\cdot05} = 0{\cdot}26547$ against the accurate value of 0·26534. For the 1 % point $t_0 = 2{\cdot}2338$, $z_{0\cdot01} = 0{\cdot}3754$ against the accurate value of 0·3746. When $n_1 = n_2 = 24$, $z_{0\cdot05} = 0{\cdot}3423$ against the accurate value of 0·3425.

The type III approximation to the F-distribution is of some interest since for n_1 moderate and n_2 large, n_1F tends to be distributed as χ^2 with n_1 degrees of freedom. Since

$$\text{Mean } F = \bar{F} = \frac{n_2}{n_2-2}; \quad \sigma_F = \frac{n_2}{n_2-2}\sqrt{\frac{2(n_1+n_2-2)}{n_1(n_2-4)}},$$

$$\alpha_{3:F} = \frac{4(2n_1+n_2-2)}{n_1(n_2-6)}\sqrt{\frac{n_1(n_2-4)}{2(n_1+n_2-2)}}, \quad \alpha_{3:F} = \sqrt{(\beta_{1:F})},$$

we find the 5, 1 or 0·1 % points for F by using

$$F_{0\cdot05} = \bar{F} + \sigma_F(1{\cdot}64485 + 0{\cdot}28392\alpha_{3:F} - 0{\cdot}04902\alpha_{3:F}^2),$$

$$F_{0\cdot01} = \bar{F} + \sigma_F(2{\cdot}32635 + 0{\cdot}73330\alpha_{3:F} - 0{\cdot}024957\alpha_{3:F}^2),$$

$$F_{0\cdot001} = \bar{F} + \sigma_F(3{\cdot}0903 + 1{\cdot}4190\alpha_{3:F} + 0{\cdot}05667\alpha_{3:F}^2).$$

* [Both Dr Wishart, as author, and myself as editor regret that owing to wartime preoccupation the publication of Dr Aroian's 1941 paper was overlooked. E.S.P.]

These formulae for the levels of significance of the χ^2 distribution are from a previous paper (Aroian, 1943). For $n_1 = 24$, $n_2 = 60$, $F_{0\cdot05}$ by this approximation is 1·709 compared with the accurate value of 1·700. For $n_1 = 24$, $n_2 = 100$, $F_{0\cdot05}$ by this approximation is 1·631 against the accurate value of 1·627. For $n_1 = n_2 = 100$, $F_{0\cdot05}$ by this approximation is 1·394 as compared with the accurate value of 1·392. While these results are not too poor, they are not so accurate as the well-known formulae of Cochran-Fisher or of E. Paulson (1942) which, for large values of n_1 and n_2, generally give 4 significant figures.

REFERENCES

AROIAN, L. A. (1941). *Ann. Math. Statist.* **12**, 429.
AROIAN, L. A. (1943). *Ann. Math. Statist.* **14**, 93.
PAULSON, E. (1942). *Ann. Math. Statist.* **13**, 233.
WISHART, J. (1947). *Biometrika*, **34**, 170.

A note on the mean deviation from the median

BY K. R. NAIR

For samples drawn from a normal universe, Godwin (1945) obtained the sampling distribution of the mean deviation when the individual deviations are measured from the sample *mean*. It is well known that the mean deviation is least when it is measured from the sample *median*.* Let us refer to them as 'mean deviation from mean' and 'mean deviation from median' respectively, and use the letters m and m' to denote their sample estimates.

The exact sampling distribution of m being now known and its probability integral tabulated, the question may well be asked what the distribution of m' is. Since $m' \leqslant m$, their expectations have the same relationship

$$E(m') \leqslant E(m). \tag{1}$$

For samples of n from a normal population with standard deviation, σ, $E(m') = f'_n\sigma$ and $E(m) = f_n\sigma$, where $f'_n \leqslant f_n$. For getting unbiased estimates of σ we should divide m' by f'_n and m by f_n. What we are now interested to know is which of the two estimates has a smaller standard error. In the case of m, it has been shown by Helmert (1876) and Fisher (1920) that

$$f_n = \sqrt{\frac{2(n-1)}{n\pi}}, \tag{2}$$

and

$$\text{S.E. of } \left(\frac{m}{f_n}\right) = \frac{\sigma}{\sqrt{n}}\sqrt{\left[\left\{\frac{\pi}{2} + \sqrt{[n(n-2)]} - n + \sin^{-1}\frac{1}{n-1}\right\}\right]}. \tag{3}$$

In the case of m', we neither know f'_n nor the standard error of (m'/f'_n) for samples of size n.

It is obvious that when n is very large, the mean and median will differ very little from one another and hence $m' \to m$ and $f'_n \to f_n$. It is interesting to note that, at the other end of the scale, namely, when $n = 2$, m and m' are identical, and equal to one-half the sample range.

To discover any real difference that may exist between the standard errors of (m/f_n) and (m'/f'_n), which is the same as determining the difference between the coefficients of variation of m and m', we must consider samples of size greater than 2.

(i) Let us take $n = 3$, and let x_1, x_2, x_3 be the observed values arranged in order of ascending magnitude. We at once find that

$$m' = \tfrac{1}{3}(x_3 - x_1). \tag{4}$$

The distribution of m' for samples of 3 is therefore derivable from that of the range. The probability integral of the range has been tabulated by Pearson & Hartley (1942) for $n = 2$ to 20. For our purpose it is necessary only to know the values of the mean range ($\bar{w}$) and the standard error of the range (σ_w)

* When n, the sample size, is an odd number, the sample median is by definition the value of the $\frac{1}{2}(n+1)$th ranked observation. When n is even, the sample median is conventionally taken as the mean of the $\frac{1}{2}n$th and $\frac{1}{2}(n+2)$th ranked values. The mean deviation from the median will have the same magnitude whatever value, between the $\frac{1}{2}n$th and $\frac{1}{2}(n+2)$th ranked values, the median takes, when n is even. No complication is therefore introduced by accepting the conventional definition of the median for even-sized samples.

for samples of 3. This can be calculated, correct to six decimal places, from certain numerical values given by Pearson (1926). Using his figures,

$$\bar{w} = 1{\cdot}692568 \times \sigma, \tag{5}$$

$$\sigma_w = 0{\cdot}888368 \times \sigma. \tag{6}$$

The value of f_n' for sample of 3 is, therefore,

$$f_3' = \tfrac{1}{3} \times 1{\cdot}692568 = 0{\cdot}56419,$$

and the standard error of m'/f_3' is $$\frac{0{\cdot}888368}{1{\cdot}692568}\sigma = 0{\cdot}52486\sigma, \tag{7}$$

correct to five decimal places.

The corresponding values for f_3 and standard error of (m/f_3) are obtained by putting $n = 3$ in equations (2) and (3) and are

$$f_3 = \sqrt{\frac{4}{3\pi}} = 0{\cdot}65147, \tag{8}$$

and $$\text{S.E. of } (m/f_3) = \frac{\sigma}{\sqrt{3}}\sqrt{\left(\frac{2\pi}{3}+\sqrt{3}-3\right)} = 0{\cdot}52486\sigma, \tag{9}$$

correct to five decimal places.

Although (9) can be evaluated to any number of decimal places, we are not in a position to bring (7) to a higher order of accuracy than five decimal places. It is very unlikely that (7) and (9) are absolutely identical, but we may safely conclude that they are practically the same.

(ii) We next come to samples of 4. If x_1, x_2, x_3, x_4 be the observations arranged in order of ascending magnitude, the mean deviation from median is given by

$$m' = \tfrac{1}{4}(x_4 + x_3 - x_2 - x_1). \tag{10}$$

The distribution of m' follows immediately from 'some order statistic distributions for samples of size 4' obtained by Walsh (1946) and is as follows:

$$p(m')\,dm' = \frac{12}{[\sqrt{(2\pi)}]^3}\,e^{-2m'^2}\left(\int_0^{4m'} e^{-\frac{1}{2}y^2}dy\right)^2 dm'. \tag{11}$$

The probability integral of m' is given by

$$P(m') = \int_0^{m'} p(m')\,dm' = \left(\int_0^{4m'} \frac{1}{\sqrt{(2\pi)}}\,e^{-\frac{1}{2}y^2}dy\right)^3 = \left(\frac{2}{\sqrt{(2\pi)}}\int_0^{2m'} e^{-\frac{1}{2}x^2}dx\right)^3. \tag{12}$$

The values of $P(m')$ given by (12) can easily be evaluated using the normal probability integral table and are given in cols. (3) and (6) of the table below, alongside corresponding values (given in cols. (2) and (5)) for the probability integral of the mean deviation (m) from the mean, for samples of 4, copied from Godwin's (1945) tables.

Table giving the probability integral of the mean deviation from (a) mean and (b) median for samples of four observations from a normal universe ($\sigma = 1$)

m (or m')	$P(m)$	$P(m')$	m (or m')	$P(m)$	$P(m')$
0·0	0·00000	0·00000	1·3	0·96758	0·97229
0·1	0·00333	0·00398	1·4	0·98229	0·98475
0·2	0·02534	0·03003	1·5	0·99073	0·99192
0·3	0·07879	0·09204	1·6	0·99534	0·99588
0·4	0·16693	0·19139	1·7	0·99775	0·99798
0·5	0·28345	0·31818	1·8	0·99895	0·99905
0·6	0·41552	0·45629	1·9	0·99953	0·99957
0·7	0·54836	0·58951	2·0	0·99980	0·99981
0·8	0·66934	0·70592	2·1	0·99992	0·99992
0·9	0·77040	0·79954	2·2	0·99997	0·99997
1·0	0·84860	0·86962	2·3	0·99999	0·99999
1·1	0·90502	0·91888	2·4	1·00000	1·00000
1·2	0·94321	0·95162			

We note that although m and m' have an infinite range from 0 to ∞, their probability integrals rapidly approach unity, this value being reached to five decimal place accuracy when $m(m') = 2{\cdot}4\sigma$. We can approximately work out the moments of the two distributions from the table above. The values of the mean and the standard deviation (applying Sheppard's correction for grouping) of m and m' so obtained are given below:

$$\left.\begin{array}{lll} \text{Mean:} & \bar{m} = 0{\cdot}690986\sigma, & \bar{m}' = 0{\cdot}663187\sigma, \\ \text{Standard deviation:} & \sigma_m = 0{\cdot}297015\sigma, & \sigma_{m'} = 0{\cdot}292979\sigma, \\ \text{Coefficient of variation:} & \sigma_m/\bar{m} = 0{\cdot}429842, & \sigma_{m'}/\bar{m}' = 0{\cdot}441775. \end{array}\right\} \quad (13)$$

The values of $\bar{m}$ and $\sigma_m/\bar{m}$ obtained from the exact formulae (2) and (3) are

$$\left.\begin{array}{r} \bar{m} = \sigma\sqrt{\left(\dfrac{3}{2\pi}\right)} = 0{\cdot}690988\sigma, \\ \sigma_m/\bar{m} = \tfrac{1}{3}\sqrt{(\tfrac{1}{2}\pi + \sin^{-1}\tfrac{1}{3} + 2\sqrt{2} - 4)} = 0{\cdot}429842, \end{array}\right\} \quad (14)$$

showing close agreement with the values given in (13) for the mean and coefficient of variation of m. We may therefore consider the mean and coefficient of variation of m', approximately evaluated in (13), to be of sufficient accuracy to warrant the conclusion that, for samples of size 4, the mean deviation from the mean leads to a more 'efficient' estimate of the population standard deviation than the mean deviation from the median. As the distribution of the latter is not known for $n > 4$, we are not in a position to say whether this conclusion holds good, in general, for all values of n.

In conclusion, it seems worth making the following point:

(*a*) if expressions for the expectation and variance of m' were available and tables of its probability integral worked out,

(*b*) if the efficiency of the m' estimate compared to the m estimate for $n > 4$ was not appreciably worse than for the case $n = 4$,

there would be strong practical grounds for using m' rather than m in view of greater simplicity in calculation. In both cases we must first arrange the observations in order of magnitude. Then if $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$, m' may be calculated from the formula

$$m' = \frac{1}{n}\{(x_{n-t+1} + x_{n-t+2} + \ldots + x_n) - (x_1 + x_2 + \ldots + x_t)\}, \quad (15)$$

where $t = \frac{1}{2}n$ or $\frac{1}{2}(n-1)$ according as n is even or odd.

For m, however, we must also calculate the arithmetic mean $\bar{x}$ and look for x_k and x_{k+1} between which $\bar{x}$ lies. Then m can be obtained from one of the three formulae

$$\left.\begin{array}{l} \dfrac{nm}{2k} = \bar{x} - \dfrac{x_1 + x_2 + \ldots + x_k}{k}, \\[2ex] \dfrac{nm}{2(n-k)} = \dfrac{x_{k+1} + \ldots + x_n}{n-k} - \bar{x}, \\[2ex] \dfrac{n^2 m}{2k(n-k)} = \dfrac{x_{k+1} + \ldots + x_n}{n-k} - \dfrac{x_1 + \ldots + x_k}{k}. \end{array}\right\} \quad (16)$$

This certainly involves a rather longer process.

It is interesting to note that m' becomes a special case of the measure of dispersion based on difference between the sums of the first and the last r observations (in order of magnitude) suggested by Jones (1946), the range, becoming another special case of the same measure, when $r = 1$.

REFERENCES

FISHER, R. A. (1920). *Mon. Not. R. Astr. Soc.* **80**, 758.

GODWIN, H. J. (1945). *Biometrika*, **33**, 254.

HELMERT, W. (1876). *Astr. Nachr.* **88**, no. 2096.

JONES, A. E. (1946). *Biometrika*, **33**, 274.

PEARSON, E. S. (1926). *Biometrika*, **18**, 173.

PEARSON, E. S. & HARTLEY, H. O. (1942). *Biometrika*, **32**, 301.

WALSH, J. E. (1946). *Ann. Math. Statist.* **17**, 246.

On the method of paired comparisons

BY P. A. P. MORAN, *Institute of Statistics, Oxford University*

M. G. Kendall & B. Babington Smith (1940) have discussed the 'method of paired comparisons' for investigating preferences. Suppose we are given n objects $A, \ldots, K$, and an observer is asked to choose between every pair. If A is preferred to B we write $A \to B$. If the observer is not completely consistent, either because of his own inefficiency or because the objects are not really capable of being ranked in respect of the quality under consideration, he may make preferences of the type $A \to B \to C \to A$, and we call this an inconsistent or circular triad. Write d for the number of circular triads in a given experiment. Then Kendall & Babington Smith show that

$$\zeta = 1 - \frac{24d}{n^3 - n} \quad (n \text{ odd})$$

$$= 1 - \frac{24d}{n^3 - 4n} \quad (n \text{ even})$$

may be regarded as a 'coefficient of consistence' and lies between 0 and 1, being capable of attaining both these limits.

Now suppose that each comparison is made at random so that there are equal chances that $A \to B$ and $B \to A$. The distribution of d is then of interest. They calculate this distribution exactly for $n = 2, \ldots, 7$ and conjecture that its moments are given by

$$\mu_1' = \frac{1}{4}\binom{n}{3},$$

$$\mu_2 = \frac{3}{16}\binom{n}{3},$$

$$\mu_3 = -\frac{3}{32}\binom{n}{3}(n-4),$$

$$\mu_4 = \frac{3}{256}\binom{n}{3}\left\{9\binom{n-3}{3} + 39\binom{n-3}{2} + 9\binom{n-3}{1} + 7\right\},$$

these being polynomials in n which agree with their numerical calculations for $n = 2, \ldots, 7$. They also conjecture that the distribution tends to normality when n increases. In the present note we prove these statements.

Let the objects be numbered from 1 to n. Write $P_{ijk} = 1$ if the triad (i, j, k) is circular, and $P_{ijk} = 0$ if it is not. Then $d = \Sigma P_{ijk}$, the sum being taken over all such triads. Now by enumerating the various cases we see that $E(P_{ijk}) = \frac{1}{4}$ and so $\mu_1'(d) = E(\Sigma P_{ijk}) = \frac{1}{4}\binom{n}{3}$. Now consider $\mu_2'(d) = E[(\Sigma P_{ijk})^2]$. Consider the types of terms which results when we expand this. In the first place we have $\binom{n}{3}$ terms typified by P_{123}^2, and these contribute $\frac{1}{4}\binom{n}{3}$ to $\mu_2'(d)$. Similarly, we have terms typified by $P_{123}P_{145}$, $P_{123}P_{124}$ and $P_{123}P_{456}$, and the number of these are respectively $\frac{3}{2}\binom{n}{3}(n-3)(n-4)$, $3\binom{n}{3}(n-3)$ and $\binom{n}{3}\binom{n-3}{3}$, whilst their expectations are each $\frac{1}{16}$. It follows that

$$\mu_2'(d) = \frac{1}{16}\binom{n}{3}\left\{\binom{n}{3} + 3\right\},$$

and so

$$\mu_2(d) = \frac{3}{16}\binom{n}{3}.$$

The calculation of $\mu_3'(d)$ is a good deal more complicated. $\mu_3'(d) = E((\Sigma P_{ijk})^3)$, and on expanding we get 16 types of terms, typified by

$$\begin{array}{llll} P^2_{123}, & P^2_{123}P_{145}, & P^2_{123}P_{124}, & P^2_{123}P_{456}, \\ P_{123}P_{145}P_{167}, & P_{123}P_{124}P_{125}, & P_{123}P_{145}P_{456}, & P_{123}P_{145}P_{146}, \\ P_{123}P_{124}P_{145}, & P_{123}P_{234}P_{134}, & P_{123}P_{345}P_{567}, & P_{123}P_{245}P_{346}, \\ P_{123}P_{145}P_{678}, & P_{123}P_{124}P_{567}, & P_{123}P_{456}P_{789}, & P_{123}P_{345}P_{245}. \end{array}$$

After some calculation we find the sum of the contributions of these to be

$$\frac{1}{2304}\binom{n}{3}\{n^6-6n^5+13n^4+42n^3-158n^2-108n+864\}.$$

Reducing to the mean we get $\mu_3(d) = -\dfrac{3}{32}\dbinom{n}{3}(n-4)$.

The calculation of $\mu_4'(d)$ is a great deal more complicated, there being 85 terms which are not zero; we finally obtain, after lengthy calculations,

$$\mu_4' = \frac{1}{55296}\binom{n}{3}\{n^9-9n^8+33n^7+45n^6-582n^5+504n^4+5732n^3-10692n^2-30024n+80352\},$$

and so
$$\mu_4 = \frac{1}{55296}\binom{n}{3}\{972n^3+972n^2-36936n+80352\},$$

which reduces to the conjectured result.

We now prove that the distribution tends to normality. To do this, it is sufficient (Kendall, 1943, p. 110) to prove that

$$\frac{\mu_{2m}}{\mu_2^m}\to\frac{(2m)!}{2^m m!},\quad \frac{\mu_{2m+1}}{\mu_2^{\frac{1}{2}(2m+1)}}\to 0,\quad \text{for}\quad m=1,2,\ldots.$$

Consider the second of these first. Write $Q_{ijk} = P_{ijk}-\frac{1}{4}$. Then

$$\mu_{2m+1}(d) = E[(\Sigma Q_{ijk})^{2m+1}].$$

It is clear that for any given m we could calculate $\mu_{2m+1}(d)$ given sufficient labour, by expanding this and considering the expectation of each type of term and calculating the number of times it occurs, which will be a polynomial in n. Now consider the various types of terms in the expansion of $(\Sigma Q_{ijk})^{2m+1}$. We classify these terms according to whether the Q's have common suffixes. Let $Q_{ijk}Q_{lmn}\ldots Q_{pqr}$ be a typical product in the expansion. If this can be separated into p groups of products of Q's such that different groups have no common suffixes whilst within each group the triads are connected to each other by having common points, we shall say such a product 'contains p groups'. Moreover, the number of times such a term occurs will be a polynomial in n whose order is equal to the number of distinct suffixes occurring in the product. If in a group a suffix only appears once, the inconsistency of the triad containing it is unaffected by the remainder of the group and the expectation of the product of Q's in that group will be zero. It follows that in all those terms which contribute something non-negative to μ_{2m+1}, none of the groups can contain a suffix which appears only once. Therefore, since all terms which contain more than m groups will have at least one group consisting of a single Q, the expectation of such terms will be zero. It follows that μ_{2m+1} is a polynomial in n, of degree $3m+1 = \left[\text{integral part of } \frac{3}{2}(2m+1)\right]$ at most, whose coefficients depend on m only. But $\mu_2(d)$ is of degree 3 in n and so

$$\frac{\mu_{2m+1}}{\mu_2^{\frac{1}{2}(m+1)}} = O\left(\frac{n^{3m+1}}{n^{3m+\frac{3}{2}}}\right) = O(n^{-\frac{1}{2}}).$$

Now consider μ_{2m}. This is a polynomial in n, and our aim is to find the order and coefficient of the term of largest order. In the first place we need only consider terms with m or less groups, for if a term has more than m groups, one at least will consist of a single Q and the expectation of the term will be zero. Moreover as before, in each term, the suffixes in each group must each occur at least twice in that group. The number of times each type of term occurs will be a polynomial in n of order equal to the total number of distinct suffixes in that term. As we shall show the leading term in $\mu_{2m}(d)$ to be of order $3m$, we can neglect terms whose frequency is less than this and therefore we can neglect all terms in which a suffix appears more than twice. Now consider a term with fewer than m groups and therefore containing a group of

order greater than two in the Q's. As no suffix can occur more than twice, no Q can occur more than once. Consider any Q_{ijk}, say, of this group. Then either the suffixes i, j, k are common to three other triads or one, i, say, is common to another triad and j, k common to a third. In either case evaluation of the expectation shows it to be zero. We can therefore restrict our attention to the case where there are m groups each containing two triads. Such groups can only be of the form Q^2_{123}, $Q_{123}Q_{124}$, $Q_{123}Q_{145}$ and the expectations of the two latter are zero whilst the expectation of Q^2_{123} is

$$\tfrac{1}{4}(1-\tfrac{1}{4})^2+\tfrac{3}{4}(\tfrac{1}{4})^2 = \tfrac{3}{32}.$$

The number of groups is m and the number of ways of choosing m such distinct pairs out of $(\Sigma Q_{ijk})^{2m}$ is $\dfrac{(2m)!}{2^m m!}$ so that the leading term in μ_{2m} is

$$\frac{(2m)!}{2^m m!}(\tfrac{3}{32})^m n^{3m}$$

whilst the leading term in μ_2 is $\frac{3}{32}n^3$ and so

$$\frac{\mu_{2m}}{\mu_2^m} \to \frac{(2m)!}{2^m m!}.$$

The distribution therefore tends to normality.

REFERENCES

Kendall, M. G. & Babington Smith, B. (1940). *Biometrika*, **31**, 324.

Kendall, M. G. (1943). *The Advanced theory of Statistics*, **1**. London: Charles Griffin and Co.

Notes on the calculation of autocorrelations of linear autoregressive schemes

By M. H. QUENOUILLE

1. Bartlett (1946) has recently shown how, for a series of observations, we can test whether the observations can be adequately represented by a linear autoregressive scheme

$$u_{n+j}+a_1u_{n+j-1}+\ldots+a_ju_n = \epsilon_{n+j}, \tag{1}$$

where the a_i are known or fitted values, and ϵ_{n+j} is an error component independent of u_{n+j-1}. Bartlett's test is based on the formula

$$\operatorname{cov}(r_s, r_{s+t}) \sim \frac{1}{n-s}\sum_{i=-\infty}^{\infty}\rho_i\rho_{i+t},$$

where r_s is the estimate of the true autocorrelation ρ_s between u_i and u_{i+s}.

The purpose of the note is to demonstrate how, using generating functions, ρ_i and $\sum_{i=-\infty}^{\infty}\rho_i\rho_{i+t}$ can be calculated with the minimum of computation.

2. The method of generating functions seems to have been used by Wold (1938), who applied them to finding the variances and covariances of linear forms of finite extent in variables such as ϵ_{n+j}. We shall, however, be concerned with linear forms of infinite extent.

It can easily be shown that the solution of (1) can be written

$$u_n = \epsilon_n + b_1\epsilon_{n-1}+b_2\epsilon_{n-2}+\ldots, \tag{2}$$

where
$$(1+a_1t+\ldots+a_jt^j)^{-1} = 1+b_1t+b_2t^2+\ldots. \tag{3}$$

For example, if
$$u_{n+2}+au_{n+1}+bu_n = \epsilon_{n+2},$$

we have
$$(1+at+bt^2)^{-1} = (1-2x\cos\theta+x^2)^{-1} = 1+\frac{\sin 2\theta}{\sin\theta}x+\frac{\sin 3\theta}{\sin\theta}x^2+\ldots,$$

where
$$\cos\theta = -\tfrac{1}{2}a/\sqrt{b},\ x = t\sqrt{b},$$

and hence
$$b_l = b^{\frac{1}{2}l}\frac{\sin l\theta}{\sin\theta} = \frac{2b^{\frac{1}{2}l}}{\sqrt{(4b-a^2)}}\sin l\theta.$$

3. Using this generating function we have

$$\sigma^2 \sum_{i=-\infty}^{\infty} \rho_i t^i = \operatorname*{Lt}_{c\to 1}\left[\{1+a_1 t+\ldots+a_j t^j\}\left\{1+a_1\frac{c}{t}+\ldots+a_j\left(\frac{c}{t}\right)^j\right\}\right]^{-1}. \tag{4}$$

Now the expansion of (4) can be achieved by splitting into partial fractions and, in general, we can let $c\to 1$ before this operation is performed. Thus

$$\frac{}{(1+a_1t+\ldots+a_jt^j)(t^j+a_1t^{j-1}+\ldots+a_j)} = \frac{A_0+A_1t+\ldots+A_{j-1}t^{j-1}}{1+a_1t+\ldots+a_jt^j}+\frac{B_j+B_{j-1}t+\ldots+B_1t^{j-1}}{t^j+a_1t^{j-1}+\ldots+a_j}, \tag{5}$$

and using $\rho_i = -\rho_{-i}$, we can see that

$$A_i = -B_j/a_j \quad (i=0)$$
$$= B_i - a_i B_j/a_j \quad (i=1,\ldots j-1).$$

Thus the autocorrelations will be generated by

$$1+\frac{t}{A_0}\,\frac{B_1+B_2t+\ldots+B_jt^{j-1}}{1+a_1t+\ldots+a_jt^j}+\frac{1}{A_0t}\,\frac{B_1+B_2t^{-1}+\ldots+B_jt^{-j+1}}{1+a_1t^{-1}+\ldots+a_jt^{-j}}, \tag{6}$$

where the first term is expanded in powers of t and the second term is expanded in powers of t^{-1}.

4. The expression (6) can now be squared to give a generating function for $\sum_{i=-\infty}^{\infty} \rho_i\rho_{i+t}$. It will be necessary to split

$$\frac{t(B_1+B_2t+\ldots+B_jt^{j-1})(B_1t^{j-1}+B_2t^{j-2}+\ldots+B_j)}{(1+a_1t+\ldots+a_jt^j)(t^j+a_1t^{j-1}+\ldots+a_j)}$$

into partial fractions, but the labour will be reduced since the matrix of the coefficients of the equations in B_i will be unaltered.

5. To illustrate the method, we can consider Kendall's series 1, which was used by Bartlett in his example.

The autoregressive scheme for this series is

$$u_{n+2}-1\cdot 1u_{n+1}+0\cdot 5u_n = \epsilon_{n+2},$$

so that $$\sigma^2\sum_{i=-\infty}^{\infty}\rho_it^i = \frac{t^2}{(1-1\cdot 1t+0\cdot 5t^2)(t^2-1\cdot 1t+0\cdot 5)} = \frac{2B_2+(B_1+2\cdot 2B_2)t}{1-1\cdot 1t+0\cdot 5t^2}+\frac{B_2+B_1t}{t^2-1\cdot 1t+0\cdot 5},$$

where

$$\begin{bmatrix}B_1\\B_2\end{bmatrix} = \begin{bmatrix}-3\cdot 7692 & 2\cdot 1154\\ -2\cdot 1154 & -1\cdot 4423\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix}$$
$$\begin{bmatrix}2\cdot 1154\\ -1\cdot 4423\end{bmatrix}.$$

Thus $$\sigma^2 = 2\cdot 8846,$$

and $$\sum_{i=-\infty}^{\infty}\rho_it^i = 1+t\frac{0\cdot 7333-0\cdot 5t}{1-1\cdot 1t+0\cdot 5t^2}+\frac{1}{t}\,\frac{0\cdot 7333-0\cdot 5t^{-1}}{1-1\cdot 1t^{-1}+0\cdot 5t^{-2}}, \tag{7}$$

so that $$\rho_i-1\cdot 1\rho_{i-1}+0\cdot 5\rho_{i-2} = 0 \quad (i>0). \tag{8}$$

If we now consider the square of the expression (7) we have a product term

$$\frac{2t(0\cdot 7333-0\cdot 5t)(0\cdot 7333t-0\cdot 5)}{(1-1\cdot 1t+0\cdot 5t^2)(t^2-1\cdot 1t+0\cdot 5)} = \frac{-2B_2+(B_1+2\cdot 2B_2)t}{1-1\cdot 1t+0\cdot 5t^2}+\frac{B_2+B_1t}{t^2-1\cdot 1t+0\cdot 5},$$

where

$$\begin{bmatrix}B_1\\B_2\end{bmatrix} = \begin{bmatrix}-3\cdot 7692 & 2\cdot 1154\\ -2\cdot 1154 & -1\cdot 4423\end{bmatrix}\begin{bmatrix}-0\cdot 7333\\ 1\cdot 5754\end{bmatrix}$$
$$= \begin{bmatrix}0\cdot 5686\\ -0\cdot 7210\end{bmatrix},$$

and, if we write
$$P_t = \sum_{i=-\infty}^{\infty} \rho_i \rho_{i+t} \Big/ \sum_{i=-\infty}^{\infty} \rho_i^2,$$

$$\begin{aligned}\sum_{i=-\infty}^{\infty} \rho_i^2 \sum_{i=-\infty}^{\infty} P_i t^i &= 1 + 2t\frac{0\cdot7333-0\cdot5t}{1-1\cdot1t+0\cdot5t^2} + t^2\left(\frac{0\cdot7333-0\cdot5t}{1-1\cdot1t+0\cdot5t^2}\right)^2 \\ &\quad + 1\cdot4420 + t\frac{0\cdot5686-0\cdot7210t}{1-1\cdot1t+0\cdot5t^2} + \text{terms in } t^{-1} \\ &= 2\cdot4420 + t\frac{2\cdot0352-1\cdot7210}{1-1\cdot1t+0\cdot5t^2} + t^2\frac{0\cdot5377-0\cdot7333t+0\cdot25t^2}{(1-1\cdot1t+0\cdot5t^2)} \\ &\quad + \text{terms in } t^{-1}. \end{aligned} \tag{9}$$

From this we have $\sum_{i=-\infty}^{\infty} \rho_i^2 = 2\cdot4420$ and the 'correlations' P_i of the correlations are 0·8334, 0·4321, 0·0006, ... Successive terms may be calculated using the relation

$$P_i - 2\cdot21 P_{i-1} + 2\cdot2 P_{i-2} - 1\cdot1 P_{i-3} + 0\cdot25 P_{i-4} = 0 \quad (i > 0). \tag{10}$$

The calculation of $\sum_{i=-\infty}^{\infty} P_i^2$, suggested by Bartlett, can also be made by this method, but it is more arduous, and the first few terms will give a good approximation.

7. The same method can be used to calculate the appropriate number of degrees of freedom for testing the correlation between two linear autoregressive schemes.

In general, if $E(u_i u_j) = \rho_{ij}\sigma^2$, $E(v_i v_j) = \rho'_{ij}\sigma'^2$ and $r = \dfrac{\sum_{i=1}^{n} u_i v_i}{\sqrt{\left(\sum_{i=1}^{n} u_i^2 \sum_{i=1}^{n} v_i^2\right)}}$,

then
$$\operatorname{var} r \sim \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} \rho_{ij}\rho'_{ij}}{\sum_{i=1}^{n} \rho_{ii} \sum_{i=1}^{n} \rho'_{ii}}.$$

For linear autoregressive schemes, $\rho_{ij} = \rho_{i-j}$, $\rho'_{ij} = \rho'_{i-j}$, and thus

$$\begin{aligned}\operatorname{var} r &\sim \frac{n + (n-1)\rho_1\rho'_1 + \ldots + \rho_{n-1}\rho'_{n-1}}{n^2} \\ &\sim \sum_{i=-\infty}^{\infty} \rho_i \rho'_i / n.\end{aligned}$$

Thus, provided n is large, r can be tested with $n \Big/ \sum_{i=-\infty}^{\infty} \rho_i\rho'_i$ degrees of freedom, and the calculation of $\sum_{i=-\infty}^{\infty} \rho_i\rho'_i$ can be made by the above method.

8. Finally, it is worth noting that, for autoregressive schemes involving m observables, it is possible to extend this method by the use of m parameters to calculate the correlations within and between the observables, provided that adequate estimates of the coefficients of the equations are available. In practice, however, the procedure will often be reversed, and estimates of the coefficients of the autoregressive schemes will be obtained by equating the theoretical and observed correlations.

REFERENCES

Bartlett, M. S. (1946). *J.R. statist. Soc. Suppl.* **8**, 27.

Wold, H. (1938). *A Study in the Analysis of Stationary Time-series.* Uppsala.

Approximate formulae for the percentage points of the incomplete beta function and of the χ^2 distribution

By D. HALTON THOMSON

Valuable 'Tables of percentage points of the Incomplete Beta Function' have been published in *Biometrika* (Thompson, 1941*a*) giving numerical values of percentage points at various probability levels between $P = 0{\cdot}995$ and $P = 0{\cdot}005$ for degrees of freedom $\nu_1 = 2q$ and $\nu_2 = 2p$ ranging up to 120, and with an accuracy of five significant figures. In the same volume, a 'Table of the percentage points of the χ^2 distribution' was also published (Thompson, 1941*b*) for values at the same probability levels and degrees of freedom ranging up to $\nu = 100$, and with an accuracy of six significant figures, thus supplementing the table of that function originally due to R. A. Fisher (Fisher & Yates, 1938).

Cases arise in practice where the tails of the frequency distribution of a large population are of special interest, thus involving (in the case of the beta function) values of $2p$ larger than 120, with a small $2q$, or vice versa. Harmonic interpolation between 120 and infinity, however, leads to substantial errors, as is found when the values of the percentage points x are expressed in terms of their tail values (x or $1-x<0{\cdot}5$). This Note shows that close approximations to such extreme values may be determined by using the χ^2 table as an auxiliary table to extend the Beta Function Tables in conjunction with certain simple alternative formulae. Comparisons within the range of the published Beta Function Tables are made indicating the degree of accuracy within that range. The accuracy of these formulae beyond that range increases rapidly with increasing $2p$ and decreasing $2q$ (and vice versa), so that they can be applied with confidence under such conditions.

The 'normalized' form of the Incomplete Beta Function, in the usual notation, is

$$P = I_x(p,q) = \frac{\Gamma(p+q)}{\Gamma(p)\,\Gamma(q)}\int_0^x x^{p-1}(1-x)^{q-1}\,dx, \tag{1}$$

in which, for a given P, $1-x(q,p)$ in the tables denotes the upper percentage point and $x(p,q)$ the lower percentage point.

It is known that, when p is large and q is small compared with p, this form tends towards the Incomplete Gamma Function

$$P = \frac{p^q}{\Gamma(q)}\int_0^t e^{-pt}t^{q-1}\,dt,$$

where $x(p,q) = e^{-t}$. This in turn may be transformed to the χ^2 distribution by putting $pt = [\chi^2_\nu(P)]/2$. For a given large p and small q, therefore, the percentage point in terms of χ^2 is given approximately by

$$x(2p,2q) \simeq \exp\left[-\frac{\chi^2_{2q}(P)}{2p}\right], \tag{2}$$

where $2q = \nu$ in the χ^2 table. This expression gives the exact value of x, when $2q = 2$, but for larger $2q$ the error, which is consistently negative, increases rapidly with increasing $2q$ unless $2p$ is very large—much larger than 120. It is, therefore, of limited practical use. The following modifications were in consequence evolved.

Approximation A

Consider the constant of integration in the original form (1) which, when expanded, is

$$\frac{(p+q-1)(p+q-2)\dots(p+1)\,p}{\Gamma(q)}.$$

Let the terms $q-1, q-2, \dots, 1, 0$ be averaged; the constant as a first approximation then becomes

$$\frac{\{p+\tfrac{1}{2}(q-1)\}^q}{\Gamma(q)}.$$

The numerator suggests that a more accurate approximation for x would be obtained by substituting $p+\tfrac{1}{2}(q-1)$ in place of p in (2), thus leading to

$$x(2p,2q) \simeq \exp\left[-\frac{\chi^2_{2q}(P)}{2p+q-1}\right]. \tag{A}$$

A comparison of the approximate values of x obtained from (A) with the exact values in the Beta Function Tables, for all probability levels between $P = 0{\cdot}995$ and $P = 0{\cdot}005$, shows that:

(*a*) The error is consistently positive, but much smaller than the negative error in (2); in other words the latter is slightly over-corrected.

(*b*) For a given p/q and varying P, the error is nearly constant; it is smallest at $P = 0{\cdot}995$ and increases gradually in the direction of $P = 0{\cdot}005$.

(*c*) For a given P, the error decreases rapidly with increasing $2p$ and/or decreasing $2q$.

(*d*) Provided that p/q is larger than 4, the value of x is within 0·5 % of the exact tail value; if p/q is larger than 10, the error is within 0·1 % of that value.

Approximation B

The exponent in (A) may be written

$$\frac{\chi^2_{2q}(P)}{2p+q-1} = \frac{\chi^2_{2q}(P)}{2q}\,\frac{2q}{2p+q-1}$$

$$= \frac{\chi^2_{2q}(P)}{2q}\left[\frac{1}{2\left(\dfrac{2p+2q-1}{2q}\right)-1}\right].$$

The factor in square brackets is equivalent to the first term in the known expansion of the form

$$\log\left(\frac{n-1}{n}\right) = -2\left\{\frac{1}{2n-1}+\frac{1}{3(2n-1)^3}+\ldots\right\},$$

where $n = (2p+2q-1)/2q$, which converges rapidly when n is large; i.e. when $2p$ is large compared with $2q$. The above exponent may therefore be written

$$-\frac{\chi^2_{2q}(P)}{2q}\log\left(\frac{2p-1}{2p+2q-1}\right),$$

which, when inserted in (A), leads to
$$x(2p, 2q) \simeq \left(\frac{2p-1}{2p+2q-1}\right)^k, \qquad \text{(B)}$$
where $k = [\chi^2_{2q}(P)]/(2q)$.

A similar comparison with the Beta Function Tables, for the same range of probability levels, shows that:

(*a*) Approximation (B) gives generally more accurate values than (A), except when $2q$ is very small, in which case they are nearly identical.

(*b*) For a given p/q and varying P, the error is negligible in the vicinity of $P = 0{\cdot}25$; it increases negatively in the direction of $P = 0{\cdot}995$, and positively in the direction of $P = 0{\cdot}005$, the largest errors occurring at this level.

(*c*) For a given P, the error decreases rapidly with increasing $2p$ and/or decreasing $2q$.

(*d*) Provided that $(2p)^3/(2q)^2$ is larger than about 150, the values of x are within 0·5 % of the exact tail value: this implies that if $2p$ is larger than about 150, this degree of accuracy is attained even when p/q is as low as unity. If $(2p)^3/(2q)^2$ is larger than about 2000, the error is within 0·1 % of the exact value, which implies that if $2p$ is larger than about 120, this degree of accuracy is attained when p/q is as low as 4.

It will be observed that, when $2q = 2$, the formula does not revert exactly to (2), as is required by theory; but, unless $2p$ is also quite small, the error in the computed value of x is negligible.

The expansion of (B) leads to
$$x(2p, 2q) \simeq e^{-v} - s(1+\tfrac{1}{2}s)\,v + sv^2,$$
where
$$v = \frac{\chi^2_{2q}(P)}{2p+2q-1} \quad\text{and}\quad s = \frac{q}{2p+2q-1},$$
thus demonstrating its analogies with Campbell's formula (C) below.

Adaptation of Campbell's formula

In a book concerned primarily with quality control, Simon (1941) quotes (without the proof) a formula, due to Campbell (1923), designed to determine the average number of defectives in a sample of n, starting from the known average number in an infinite sample. It is a particular application of the general problem now under consideration, namely, the approximate determination of the percentage points of

the Beta Function, starting from the corresponding known values for the χ^2 form of the Poisson exponential binomial summation. It is given in the following form:

$$\frac{a(c,n,P)-a(c,\infty,P)}{a(c,\infty,P)} = An^{-1}+\tfrac{1}{12}[14A^2+(3a+2)A+a]n^{-2}+\ldots, \tag{3}$$

where $a(c,n,P)$ = average number of defectives in which P is the probability of at least c defectives in a sample of n, $a(c,\infty,P)$ = average number of defectives in an infinite sample, $A = \frac{1}{2}(c-a-1)$, in which $a = a(c,\infty,P)$. (Simon quotes $a = (a,\infty,P)$, which is an evident misprint.)

If G denotes the value given by the formula, then

$$a(c,n,P) = a(c,\infty,P)(1+G),$$

so that $1+G$ is the factor by which the average number of defectives in an infinite sample must be multiplied to give that in a sample of n.

The change from Campbell's notation to the more familiar general notation is given by

$$a(c,n,P) = \{1-x(2p,2q)\}\,n, \quad a = a(c,\infty,P) = [\chi^2_{2c}(P)]/2,$$

where $$n = p+q-1, \text{ and } c = q.$$

Let $$u = a/n \quad\text{and}\quad r = (c-1)/(2n),$$

then $$A = n(r-u/2).$$

By inserting this notation in (3) and rearranging the terms, the formula leads to

$$x(2p,2q) \cong 1-\left\{1+r\left(1+\tfrac{7}{6}r+\frac{1}{6n}\right)\right\}u+(1+\tfrac{11}{6}r)\frac{u^2}{2}-\tfrac{1}{6}u^3+\ldots,$$

which expression includes the first four terms in the expansion of e^{-u}.

Hence, for the determination of the percentage points $x(2p,2q)$, Campbell's formula may, in effect, be re-written as

$$x(2p,2q) \cong e^{-u}-r\left(1+\tfrac{7}{6}r+\frac{1}{6n}\right)u+\tfrac{11}{12}ru^2, \tag{C}$$

where $$u = \frac{\chi^2_{2c}(P)}{2(p+q-1)} \quad\text{and}\quad r = \frac{q-1}{2(p+q-1)}.$$

For large $2p$ and small $2q$, the last two terms become negligible, in which case it reduces to

$$x(2p,2q) \cong e^{-u}-r(1+\tfrac{7}{6}r)\,u. \tag{C$'$}$$

Cochran's approximation

Cochran (1940), extending a method of Fisher's (1925), has introduced a useful approximation for the percentage points of the Incomplete Beta Function, *when both p and q are large*, his method being to determine a sufficiently accurate value of z, as used in Fisher's z-transformation.

If y is the normal deviate at probability level P, then for a given pair of arguments $2p$, $2q$, the following are first calculated, using Hartley's (1941) notation:

$$\lambda = \tfrac{1}{6}(y^2+3), \quad A = \frac{8pq}{2p+2q},$$

$$z = \frac{y}{\sqrt{(A-\lambda)}} + \frac{(\lambda-\frac{5}{6})(A-2p)}{pA}.$$

Hence, by Fisher's transformation, $$x(2p,2q) \cong \frac{2p}{2p+2qe^{2z}}. \tag{D}$$

Comparison of formulae

Table 1 compares the various formulae for upper percentage points at an extreme probability level ($P = 0{\cdot}995$). Table 2 indicates their relative accuracy on a common basis, namely, as a percentage of the exact value of x or $1-x$, whichever is the smaller, so that the deviations from the exact values, when x or $1-x$ approach zero, are duly emphasized. For intermediate probability levels, the percentages lie between the tabulated extremes. It will be noted that in the case of approximation B the errors pass through zero near the mid-range of P; in the cases of A and C the errors are positive for all values of P.

The general conclusions from these tables and other comparisons are that, for a given probability level P:

(*a*) When $p/q >$ about 6, approximations A, B and C have about the same degree of accuracy, so that the simpler, A or B, have the advantage.

(*b*) In the range $6 > p/q > 4$, there is little to choose between B and C; but B is the simpler.

(*c*) When $p/q < 4$ and the distribution approaches symmetry, D gives the best results, provided that $2p$ and $2q$ are moderately large, say > 50. It may be, however, that B in this range will be sufficiently accurate for many purposes; if $p/q > 2$, the maximum error of x is about 2 units in the third decimal place.

Table 1. *Comparison of approximate formulae at a given probability level*

$P = 0{\cdot}995$

$2p$	$2q$	$x(2p, 2q)$ A	B	C (Campbell)	D (Cochran)	Exact
120	2	$0{\cdot}9^{4}16461$ *Nil*	$0{\cdot}9^{4}16459$ *$-0{\cdot}0^{4}00002$*	$0{\cdot}9^{4}16461$ *Nil*	0·999862 *−0·000054*	$0{\cdot}9^{4}16461$ —
	4	0·9982908 *+0·0000001*	0·9982906 *−0·0000001*	0·9982907 *Nil*	0·997926 *−0·000365*	0·9982907 —
	10	0·982764 *+0·000005*	0·982755 *−0·000004*	0·982760 *+0·000001*	0·982076 *−0·000683*	0·982759 —
	20	0·944002 *+0·000072*	0·943893 *−0·000037*	0·943941 *+0·000011*	0·943366 *−0·000564*	0·943930 —
	30	0·902230 *+0·000280*	0·901839 *−0·000111*	0·902000 *+0·000050*	0·901551 *−0·000399*	0·901950 —
	40	0·86160 *+0·00067*	0·86070 *−0·00023*	0·86106 *+0·00013*	0·86066 *−0·00027*	0·86093 —
	60	0·78782 *+0·00203*	0·78522 *−0·00057*	0·78621 *+0·00042*	0·78568 *−0·00011*	0·78579 —
	120	0·62598 *+0·00978*	0·61430 *−0·00190*	0·61855 *+0·00235*	0·61620 *Nil*	0·61620 —

N.B. The figures in italics are the differences between the approximate and exact values.

Table 2. *Relative accuracy of approximate formulae*

Error of $x(2p, 2q)$ expressed as a percentage of the smaller exact tail value (x or $1-x < 0{\cdot}5$)

$2p$	$2q$		A			B			C (Campbell)			D (Cochran)		
		P / p/q	0·995	0·500	0·005	0·995	0·500	0·005	0·995	0·500	0·005	0·995	0·500	0·005
			%	%	%	%	%	%	%	%	%	%	%	%
120	12	10	*	*	*	*	*	*	*	*	*	−2·8	−0·1	+0·3
	20	6	+0·1	+0·2	+0·2	−0·1	*	+0·1	*	*	+0·1	−1·0	*	+0·1
	40	3	+0·5	+0·6	+0·7	−0·2	*	+0·2	+0·1	+0·1	+0·3	−0·2	*	*
	60	2	+0·9	+1·1	+1·2	−0·3	*	+0·2	+0·2	+0·1	+0·5	−0·1	*	*
	120	1	+2·5	+2·7	+4·4	−0·5	*	+0·7	+0·6	+0·5	+1·7	*	*	*
30	3	10	*	*	+0·1	*	*	*	*	*	+0·1	−30·0	−1·4	−7·1
	5	6	+0·1	+0·1	+0·3	−0·1	−0·1	+0·1	*	*	+0·4	−11·1	−0·5	−1·8
	10	3	+0·4	+0·5	+0·8	−0·3	−0·1	+0·3	+0·1	+0·1	+1·3	−2·2	−0·1	−0·5
	15	2	+0·8	+1·0	+1·9	−0·5	−0·1	+0·6	+0·3	+0·1	+2·8	−0·7	*	−0·3
	30	1	+2·4	+2·7	+7·1	−1·0	−0·1	+1·8	+1·0	+0·5	+8·2	*	*	*

* Error smaller than ±0·05 %.

Wilson-Hilferty approximation for χ^2-adjustment

This formula (Wilson & Hilferty, 1931) for the percentage points of the χ^2 distribution is

$$\chi^2_\nu(P) = \nu\left\{1 - \frac{2}{9\nu} + y_P\sqrt{\frac{2}{9\nu}}\right\}^3,$$

where ν represents the degrees of freedom, and y_P the standardized normal deviate corresponding to probability level P. A table has been published in *Biometrika* (Merrington, 1941), comparing the approximations derived from this formula with the exact values, at various probability levels between $P = 0{\cdot}995$ and $P = 0{\cdot}005$. It shows the remarkable accuracy of the formula, the maximum errors varying from about $\pm 0{\cdot}04$, when $\nu = 30$, to about $\pm 0{\cdot}024$, when $\nu = 100$.

When these errors were plotted against the exact values on logarithmic paper, it was observed that for a given probability level, they varied inversely with $\sqrt{\nu}$ very closely. It follows that this square root relation may be used to adjust the Wilson-Hilferty formula, bringing the values computed therefrom still nearer to the exact values.

If the difference (at $\nu = 30$) between the Wilson-Hilferty value and the exact value, when multiplied by $\sqrt{(100/30)}$, is treated as a coefficient C (which may be positive or negative), the required adjustment for any value of ν is given by

$$\text{Adjustment} = C/\sqrt{\nu}.$$

For various probability levels P, the values of C are given in the following table:

P	C	P	C
0·995	+0·233	0·250	+0·039
0·990	+0·157	0·100	+0·056
0·975	+0·067	0·050	+0·035
0·950	+0·011	0·025	−0·015
0·900	−0·029	0·010	−0·120
0·750	−0·046	0·005	−0·227
0·500	−0·013	—	—

A test against the Merrington Table shows that this adjustment leads to values of χ^2, between $\nu = 30$ and $\nu = 100$ at all probability levels with an accuracy of $\pm 0{\cdot}001$, i.e. to four or five significant figures. Since the Wilson-Hilferty approximation assumes a normal distribution about $1 - 2/(9\nu)$, which tends to unity as ν increases to infinity, and since the adjustment tends to zero under those conditions, it follows that the latter may also be safely applied for an indefinitely large ν.

It should be added that an adjustment on similar principles is not applicable to the Fisher approximation for χ^2.

REFERENCES

Campbell, G. A. (1923). *Bell Syst. Tech. J.* January.

Cochran, W. G. (1940). Note on an approximate formula for significance levels of z. *Ann. Math. Statist.* **11**, 93.

Fisher, R. A. (1925). *Statistical Methods for Research Workers.* Edinburgh: Oliver and Boyd. 1st edition.

Fisher, R. A. & Yates, F. (1938), *Statistical Tables for Biological Agricultural and Medical Research.* Edinburgh: Oliver and Boyd.

Hartley, H. O. (1941). Tables of percentage points of the Incomplete Beta Function. Methods of interpolation. *Biometrika*, **32**, 166.

Merrington, M. (1941). Numerical approximations to the percentage points of the χ^2 distribution. *Biometrika*, **32**, 200.

Simon, L. E. (1941). *An Engineer's Manual of Statistical Methods*, p. 185. New York: John Wiley and Sons, Inc.

Thompson, C. M. (1941*a*, *b*). Tables of percentage points of the Incomplete Beta Function, *Biometrika*, **32**, 151.

Thompson, C. M. (1941). Table of percentage points of the χ^2 distribution. *Biometrika*, **32**, 187.

Wilson, E. B. & Hilferty, M. M. (1931). The distribution of chi-square. *Proc. Nat. Acad. Sci. Wash.* **17**, 684.

REVIEWS

A First Course in Mathematical Statistics. By C. E. WEATHERBURN. Cambridge University Press. Price 15*s*.

An outstanding feature of the present statistical time is the number of text-books which are being written, and each one from a slightly different point of view. It is this which makes statistical theory interesting to study, for there can be no rigid approach to a subject which is used and expounded by so many and diverse persons. Professor Weatherburn has taken a rather formal mathematical exposition of the subject, and mathematical students will find his book both interesting and profitable to read. Numerical examples are given for the reader to apply the appropriate mathematical technique. It is possible that these would have been of greater utility if they had contained the material in its crude state, and had not been streamlined so that the application of the technique is immediately obvious, but nevertheless many new examples are there.

I am not sure whether this book will be entirely useful to students of other subjects than mathematics. While the mathematical analysis is undoubtedly clear it is possible that many will not be able to follow it in detail, and the conclusions of the analysis are not emphasized strongly. We may contrast with this Fisher's *Statistical Methods for Research Workers*, where no analysis is given, but where the relevant formulae and their interpretation are stated unmistakeably and their applications to material in its crude state set out so that the student may calculate for himself.

Probability theory is the foundation stone on which the whole of statistical theory is built. It is disappointing therefore to find that it is given somewhat perfunctory treatment in one chapter and the part it plays in (say) statistical tests of significance is not brought out and emphasized. There is a tendency nowadays in applying statistical technique to regard the 5 % and 1 % levels of significance as sacrosanct and those coming fresh to the subject should learn that custom is the only reason for their choice.

In spite of the criticisms which I make, however, I would recommend this book to students who have obtained some idea of the aims and objectives of statistical theory, and who are desirous of learning the development of the mathematical technique as well as its application. Professor Weatherburn's mathematical analysis makes pleasant reading and may well throw new light on old methods for those who have learnt the rudiments of the theory.

F. N. DAVID

Advances in Genetics, Volume 1. New York, N.Y.: Academic Press. 1947.

This is the first number of a new periodical, probably an annual, summarizing recent work in various fields. Of the nine articles, ranging from 12 to 96 pages, with mean 42·6, S.D. 7·89, and a positively skew distribution, perhaps the most interesting to European geneticists will be that on the genetics of the ciliate Protozoa, *Paramecium* and *Euplotes*. Here Sonneborn describes work almost entirely done in America, with very surprising results. Thus *Paramecium aurelia* consists of at least seven endogamous varieties, each with two exogamous mating types, which might be called sexes were it not that in *P. bursaria* one of the varieties has no less than eight mating types.

Shrode and Lush's article of the genetics of cattle gives a very condensed account of the large amount of work which has been done on the inheritance of economically important characters such as milk yield and growth rate. For example cattle biometricians have used the important concept of 'heritability', meaning the fraction of the variance of a character due to additive genetic differences. Within a herd this rarely exceeds 30 %. More space is devoted to work on the genetics of colour and the like, which is of far less economic importance, and the review of progeny testing methods is disappointingly brief. However, the bibliographical references will be useful. Similarly, Atwood's article on forage crops, though most valuable as a guide to the literature, does not give a detailed account of any of the biometric work which has been done on grasses and clovers.

Only two of the papers give data which a biometrician could immediately utilize. These are Gordon's account of polymorphism in fish populations, and Spencer's of mutations in wild *Drosophila* species, which unfortunately does not include some valuable recent Italian and Russian work. Gordon's results call for the development of methods of estimating gene frequency similar to those used with human blood groups. Spencer is mainly concerned with results, but these are often given in sufficient detail to interest biometricians, though no attempt is made to summarize Wright's fundamental statistical theory.

The other articles will be less attractive to biometricians, though it is of interest to see how statistical methods are demanded by the mere fact that the genus *Crepis*, whose evolution is reviewed by Babcock, includes 196 species, most of which have been examined cytologically, and between which 130 of the 38,220 possible crosses have been made.

The volume will be indispensable to geneticists. Biometricians certainly cannot neglect it.

J. B. S. H.

Mathematical Methods of Statistics. By H. Cramér, Princeton University Press. 1946. $6.00.

This book was written by Prof. Cramér during the war and has been published first in Sweden and then by an offset process by the Princeton University Press in the U.S.A. It is a definitive exposition of the theory of mathematical statistics as it existed in 1940 (about) and it is worth while therefore to consider its contents in some detail. Prof. Cramér has divided his exposition into three parts; the first part is purely mathematical. The theory of sets and of such Lebesgue measure as is necessary for the understanding of the second part is developed first of all. Such a development will be useful for the student of mathematical statistics coming fresh to the theory of measure in that he receives guidance as to what are the elements essential for him to understand. Chapters 11 and 12 on matrices determinants and quadratic forms and miscellaneous complements do not fit into this general scheme but have obviously been included here as part of the mathematical equipment necessary for the student. Possibly Chapter 10 on Fourier Integrals would have fitted more naturally into Part II but this is a matter of taste.

Part II begins with a formal development of the theory of probability as given by the French and Russian schools of probability, and which Prof. Cramér has already given in his Cambridge tract 'Random Variables and Probability Distributions'. The treatment here seems simpler, however, than in his earlier tract and there is a more practical flavour to his exposition. This part while still purely mathematical begins to introduce distributions and ideas which are familiar to the statistician.

The title of the third part is 'Statistical Inference' and the main outline is that of small sample theory developed during the past twenty-five years. The illustrations are numerical as well as mathematical and an attempt is made to show the student the numerical applications of the processes through which his mathematical theory leads him. The treatment is not exhaustive but the student who has assimilated this part will have little difficulty in extending his knowledge by further reading.

As a textbook of mathematical statistics this book will remain unrivalled for many years to come. The mathematical exposition is clear, the development of ideas logical throughout, and the theorems are presented in a very general way. Any student of mathematics who wishes to get a picture of what statistical theory is about will be led inevitably to a study of this book. To those who wish to become statisticians it will be necessary to supplement the reading by a practical course in which the mathematical tools are tried out on numerical examples. This aspect of statistical work the book does not cover, but it is obvious that this would be the case from the title. It only remains to say to the student 'This is a good book, buy it'.

F. N. DAVID

CORRIGENDA

(*Biometrika*, **34**, 176–7)

In J. Wishart's paper on 'The cumulants of the z and of the logarithmic χ^2 and t distributions', the following correction should be made:

p. 176, 1st line of section 3: *read* '$\log|t|$' *for* '$\log t$', in two places.

p. 177, 1st line following equation (32): *read* '$\log|x|$' *for* '$\log x$'.

BIOMETRIKA

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

VOLUME XXXV

1948

ISSUED BY

THE BIOMETRIKA OFFICE, UNIVERSITY COLLEGE, LONDON

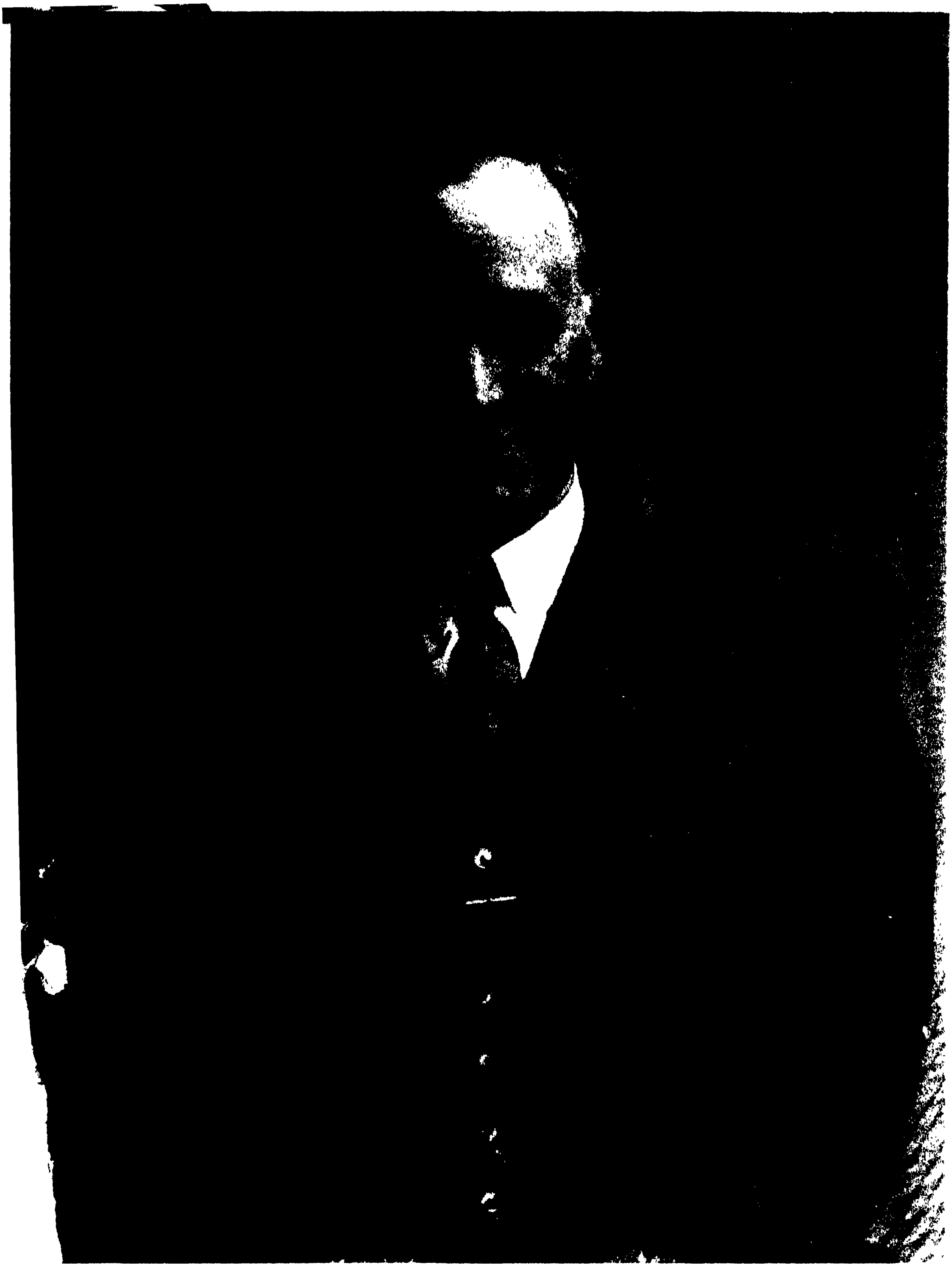
PRINTED AT THE UNIVERSITY PRESS, CAMBRIDGE

PRINTED IN GREAT BRITAIN

CONTENTS OF VOLUME XXXV

Memoirs

Miscellanea



JAMES FOWLER TOCHER

1864–1945

JAMES FOWLER TOCHER

James Fowler Tocher (1864–1945), chemist, ethnologist, biometrician, agriculturist and man of affairs, began adult life by opening a chemist's business in Peterhead more than sixty years ago. Even sixty years ago, scientific chemistry was not restricted to professors or even analysts, and certainly proprietors of businesses sometimes made money. The Southron, however, unduly influenced by jokes about Aberdonians—most of which are based on the principle, *lucus a non lucendo*—would not think Peterhead a promising venue either for scientific research or earning a competence. In fact, the business was a financial success; Tocher disposed of it in 1912 and when he passed to the analytical branch of professional chemistry, he had already made a name among chemists and biometricians (he was President of the Pharmaceutical Society of Great Britain in 1908).

Like several other men who have done important scientific work Tocher owed a good deal to a local society, the Buchan Field Club. In the south, although the name Field Club is used, e.g. the Essex Field Club, Natural History Society is perhaps a commoner designation and can include all the descriptive sciences. It is quite possible that these societies are educationally almost as valuable as classes in technical colleges even now; sixty years ago they were invaluable. The Buchan Field Club published transactions; in these appeared Tocher's earliest statistical papers, the very first in 1895.

They were concerned with the ethnology of Buchan and form the basis of the wider studies with which his name is usually associated. Among the earlier papers is one, written jointly with a frequent collaborator, Mr, afterwards Prof., James Gray, on the 'Frequency and Pigmentation Value of Surnames in East Aberdeenshire'.

More than ninety years ago, William Farr wrote an essay on the statistics of surnames which appeared in the *Sixteenth Annual Report of the Registrar General* and may have been read by Tocher because it is reprinted in the memorial volume edited by Noel Humphreys which preserves some, but by no means all, of the delightful essays by an old master. Farr used the registers of births and deaths and found in a sample of 275,405 names 32,818 different surnames or 11·9 % different surnames. In Wales the proportion was much smaller. 'The name of John Jones is a perpetual incognito in Wales, and being proclaimed at the cross of a market town would indicate no one in particular.' Farr touched on the local distributions and asked such questions as whether the present predominance of the Smiths were due to the original numerical strength of that great family, or 'to some special circumstances acting upon the ordinary laws of increase, owing to which the descendants of the hammer-men have multiplied at a greater rate than the bearers of any other name?' Did the progeny, he asked, of the tawny Browns increase faster than that of the fair-complexioned Whites?

Tocher interested himself in such problems. The data of the investigation of Gray & Tocher (printed in 1902) were pigmentation records of 14,561 school-children in East Aberdeenshire. These 14,561 children had 751 surnames, 5·2 %. A smaller proportion than in Farr's data, as one might expect, since in a school population there would be many more brothers and sisters than in the birth and death registers for two quarters. In East Aberdeenshire, Smith, scoring 203, was easily beaten by Milne with 267; third came Taylor, which is fourth on Farr's list for England and Wales. Gray & Tocher worked out the mean and standard deviation in pigmentation units of the children with different surnames. 'The

blondest surname in our list is Pirie....Next to Pirie comes the surname Wallace. This points to the conclusion that the Wallace sept sprang from ancestors with a decided blonde tendency.'

The use of the standard deviation in this paper, as well as an allusion to Karl Pearson's work on assortative mating, show that either Gray or Tocher or both were already readers of K.P. At what date Tocher made the personal acquaintance with Karl Pearson, which ripened into a warm friendship, I do not know. Five years later, in 1907,* Tocher's first full-scale biometric memoir appeared in this *Journal* (Vol. 5, pp. 298–350). The original data are printed as an appendix to the volume. It is careful biometric study of the stature, pigmentation, and craniometry of inmates of Mental Homes in Scotland. In 1908 (Vol. 6, pp. 129–235, Appendix 1–63) is an equally careful study of the pigmentation of school-children in Scotland. These researches were financed by the Henderson Trust who published in book form (Oliver & Boyd, 1924) a summary by Tocher of the results recorded in the earlier papers together with those of later anthropometric observations on samples of the civil populations of Aberdeenshire, Banffshire and Kincardineshire and on soldiers of Scottish nationality.

The use of superlatives is not a method of science—there is *no* greatest 'scientist'—but, whether any anthropometric survey of another part of the British Isles so far undertaken can take precedence over the voluntary work of Tocher supported financially not by the state but by a private trust, is a question which may fairly be asked. I think the answer is 'no'. Tocher's analysis followed the lines of teaching so many of us older people look back upon with gratitude. It may well be that, since his time, the technique of craniometry has been extended and genetic research has modified some of the perhaps rather naïve theories of inheritance popular a generation ago. But Tocher's work remains of fundamental importance; it is a pity his example did not inspire Englishmen to extend his study of mental hospital populations.

In 1925 (*Biometrika*, Vol. 17, pp. 142–58) R. Greenwood, C. M. Thompson & H. M. Woods published a memoir on heights and weights of patients and there are scattered through the literature other statistical papers, but not, I think, any large-scale investigation. But the mind-body relation still interests us all. Our remote predecessors thought there was a corporeal basis of the sanguine, melancholic, choleric and phlegmatic temperaments; psychologists and pathologists of to-day reject the bases imagined by our ancestors, but still search for *a* basis. Most people know that pulmonary tuberculosis takes a heavy toll in mental hospitals and perhaps dismiss the subject with some vague remark about the 'unfit' or, alternatively, remember that the old physicians thought that grief or emotional shock was a factor predisposing to consumption. There seems to be no doubt that a particular form of mental disorder, dementia praecox, is particularly associated with tuberculosis, so intimately that one writer has maintained that dementia praecox and pulmonary tuberculosis can, in a sense, be regarded as modifications or expressions of a common defect or diathesis. All this is, however, the merest speculation without a biometric study. Obviously to take the heights and weights of patients in institutions would only be a beginning. Tocher did not classify his data under diagnosis, perhaps he could not have done so usefully because,

* [During the summer of 1906, Tocher came south to Danby, in the Yorkshire moors, where K.P. was spending his vacation and the final arrangements for the publication of this memoir were no doubt then discussed. Tocher was also a friend of W. R. Macdonell whom he succeeded, at a rather later date, as part time lecturer in Statistics at the University of Aberdeen. *Ed.*]

although 4436 males and 3951 females sound large numbers, when they are subdivided into age- and diagnosis-groups, the classes would grow small. The records of the mental hospitals of Great Britain in the forty years since Tocher did his work would suffice. But there has not been an English Tocher.

In his later years, Tocher's deserved reputation led to official or semi-official appeals for his help in work involving statistical analysis. He was very active in the study of milk production; in tackling such problems his chemical training was of great value. A matter of obvious importance is whether poor yield is due to nature or nurture or to both. Tocher was of opinion that both factors were involved but that 'the solution to the problem of deficient milk must be the careful selection of cows with good milk records and a good milking pedigree' (see Stewart & Tocher, *J. Dairy Res.*, January 1936). Tocher did not forget, as some academic writers did, the importance of vicious circles; those who farm with insufficient capital cannot afford to buy pedigree stock and cannot afford to feed adequately what stock they have. He protested vigorously against the gross injustice of any legal presumption that because a particular sample of milk had a percentage of butter fat or a percentage of solids-not-fat below a prescribed minimum, the milk had been adulterated. It might, he thought, be quite reasonable to forbid the marketing of such milk, but certainly not right to affix a stigma to the farmer. Tocher was a good biometrician and a good practical Christian.

In attempting to give a personal impression of Tocher, a writer who also knew two of his Aberdonian friends, Charles Creighton (1847–1927) and William Bulloch (1868–1941), both of whom left classical contributions to scientific literature, is tempted to compare them. Creighton was, in the old-fashioned sense of the word, the greatest scholar of the three. His *History of Epidemics in Britain* is a classic and other of his less-known writings would have been approved by Dr Johnson. Bulloch, less familiar than Creighton with the ancient writers, had an encyclopaedic knowledge of modern pathological literature and readers of this *Journal* are likely to remember his contribution to the *Treasury of Human Inheritance*, his study of Haemophilia which, incidentally, contains many illustrations of the irony which made Bulloch a famous raconteur. I do not think Tocher was the equal of these two friends in scholarship or literary power, but—although to some who perhaps over-value formal scientific training it may sound paradoxical—I should say he was a better, a more original, *scientific* investigator than either. Epidemiologists will never forget Creighton, bacteriologists will never forget Bulloch. Biometricians, I hope, will read Tocher's papers and show their gratitude in the way he would most have appreciated, that is by completing some of the tasks to which he put his hand.

Tocher had always been a good friend of this *Journal* and he became a Trustee on the inception of the Biometrika Trust in 1936.

MAJOR GREENWOOD

BIBLIOGRAPHY OF TOCHER'S PAPERS

1889. The springs and wells of Peterhead. *Trans. Buchan Fld Cl.* **1**.

1891. Test for the detection of sesame oil in olive oil and isolation of another substance (sesamin) from sesame oil. *Pharm. J.* **21** (24 Jan.).

1893. A further note on sesamin. *Pharm. J.* **23** (25 Feb.).

1894. The production of gas from paraffin oils, and from pure members of paraffin and terpene series. *J. Soc. chem. Ind.*, *Lond.*, **13** (31 Mar. and 31 May).

1895. (With John Gray.) The ethnology of Buchan. Preliminary ethnographical observations. *Trans. Buchan Fld Cl.* **3**.

1897*a*. The ethnology of Buchan. Part II. Ethnographical survey of school-children in Buchan. *Trans. Buchan Fld Cl.* **4**.
1897*b*. Report on scales of contributions and systems of levies. British Order of Ancient Free Gardeners.
1900*a*. (With John Gray.) The physical characteristics of adults and school-children in East Aberdeenshire. *J. R. anthrop. Inst.* **30** (=N.S. 3).
1900*b*. (With John Gray.) The physical characteristics of the population of West Aberdeenshire. *J. R. anthrop. Inst.* (Reviews and Miscellanea), **30**, no. 84 (=N.S. **3**).
1900*c*. The volumetric determination of red lead. *Pharm. J.* **64** (24 Mar.).
1900*d*. Address on the Free Gardeners and the Friendly Society Movement.
1910*a*. (With John Gray.) The ethnology of Buchan. Part III. The physical characteristics of adults and school-children in East Aberdeenshire. *Trans. Buchan Fld Cl.* **6**. (Contains the same data as 1900*a*, with some additions.)
1901*b*. The volumetric determination of phenol. *Pharm. J.* **66** (23 Mar.).
1902*a*. (With John Gray.) The ethnology of Buchan. Part IV. The frequency and pigmentation value of surnames in East Aberdeenshire. *Trans. Buchan Fld Cl.* **7**.
1902*b*. (With John Gray.) The physical characteristics of the eskimo of Southampton Island, etc. *Trans. Buchan Fld Cl.* **7**.
1902*c*. Scottish universities and pharmaceutical education. *Pharm. J.* **68** (1 Mar.).
1902*d*. The oxidation and determination of uric acid and urates. *Pharm. J.* **69** (16 Aug.).
1903*a*. The characteristics of mankind. *Aberdeen Free Press* (21 Mar.).
1903*b*. A scheme of representation by territorial districts on the Pharmaceutical Council. *Pharm. J.* **71** (15 and 22 Aug.).
1905*a*. Medical inspection of schools. Address to the 31st Annual Congress of the Incorporated Sanitary Assoc. of Scotland. Inverness, Sept. 1905.
1905*b*. Anthropometric survey of the inmates of asylums in Scotland. *Henderson Tr. Rep.* **1**.
1906*a*. The activity of pepsin after brief contact with certain inorganic compounds. *Pharm. J.* **77** (28 July).
1906*b*. The detection of citrates and tartrates. *Pharm. J.* **77** (28 July).
1907*a*. The anthropometric characteristics of the inmates of asylums in Scotland. *Biometrika*, **5**, 298–350 and Suppl. (Supplement is a reprint of 1905*b* above.)
1907*b*. Anthropometry, its aims and methods. Privately printed.
1908. Pigmentation survey of school-children in Scotland. *Biometrika*, **6**, 129–235 and Suppl.
1909. Some problems of interest to pharmacists to-day. *Year Book of Pharmacy and Trans. of the British Pharmaceutical Conference.*
1910. The necessity for a national eugenic survey. *Eugen. Rev.* **2**.
1913. The criteria for purity of water—chemical and bacteriological. *Proc. Aberdeen Ass. Civil Engrs*, **13** (1912–13).
1915. (With Karl Pearson.) On criteria for the existence of differential death-rates. *Biometrika*, **11**, 159–84.
1918. Food values and costs in relation to milk production. *Scot. J. Agric.* **1**, no. 1.
1919*a*. Investigation into the milk yield of Ayrshire cows. *Trans. Highl. agric. Soc. Scot.*
1919*b*. Variations in the composition of milk. *Scot. J. Agric.* **2**, no. 3.
1919*c*. Science and the struggle for existence. *The Rotary Wheel.*
1919*d*. Report on distillery by-products. Privately printed.
1921. On the need for consolidating the law relating to the sale of food and drugs. *Trans. inc. sanit. Ass. Scot.*
1922*a*. The relationship between citric solubility of phosphates and yield of turnip crop. *Trans. Highl. agric. Soc. Scot.*
1922*b*. The citric solubility of mineral phosphates. *J. agric. Sci.* **12**, pt. 2.
1923*a*. Grass sickness in horses. (Reports by J. F. Tocher, J. W. Tocher, W. Brown and J. B. Buxton.) *Vet. Rec.* **35**.
1923*b*. Milk yields and associated factors. *Proc. World Dairy Congr.*
1924*a*. Grass sickness in horses. *Trans. Highl. agric. Soc. Scot.* (Contains same data as **1923*a*** above.)
1924*b*. Anthropometric observations on samples of the civil population of Aberdeenshire, etc. *Henderson Tr. Rep.* **2**.
1924*c*. A study of the chief physical characters of soldiers of Scottish nationality and a comparison with the physical characters of the insane population of Scotland. *Henderson Tr. Rep.* **3**.
1925*a*. Variations in the composition of milk. *Edin. Med. J.* (May).
1925*b*. Variations in the composition of milk. H.M.S.O.
1926*a*. The pharmacist, his training and vocation. *Pharm. J.* **117** (9 Oct.).

1926*b*. Errors of judgment in chemical analysis. *Analyst* (July).
1926*c*. Variations in the composition of milk. *Analyst* (Dec.).
1927*a*. Causes of variation in the proportion of butter fat in milk. *Scot. J. Agric.* **10**, no. 1.
1927*b*. Variation in the proportion of solids-not-fat in milk. *Scot. J. Agric.* **10**, no. 2.
1927*c*. Sugar in milk. *Scot. J. Agric.* **10**, no. 4.
1928. An investigation of the milk yield of dairy cows. *Biometrika*, **20** B, 105–244.
1929. The sale of milk. Suggested amendments of the law. *Scot. J. Agric.* **12**, no. 4.
1930. The study of literature. Lecture to Peterhead Literary Soc.
1931*a*. Should the law relating to the sale of milk be amended? *Trans. Yorks. agric. Soc.*
1931*b*. What is probable error? Lecture to Institute of Chemistry of Great Britain and Ireland.
1932. A New Year's Message. *Chem. & Drugg.* (2 Jan.).
1934*a*. Raw versus pasteurized milk. *Scot. J. Agric.* **17**, no. 1.
1934*b*. The services of Francis Galton and his school to physical anthropology and eugenics. *Rep. Brit. Ass.*, Section H.
1935. The proportions of certain poisonous substances in feeding stuffs and their effect on livestock. *Vet. Rec.* **47**.
1936. (With Alice Stewart.) The effect of variations in feeding on dairy cows yielding milk of poor quality. *J. Dairy Res.* **7**, no. 1.
1937. Agricultural education and research in Scotland. *Aberdeen Univ. Rev.* (Mar.).
1939*a*. Pasteurization of milk. *Trans. Yorks. agric. Soc.* no. 96.
1939*b*. Mortality among livestock due to poisons. *Vet. Rec.* **51**.
1940. The role of science. *The Second Book of Buchan.*
1941*a*. Prehistory and early history. *The Second Book of Buchan.*
1941*b*. The bracken problem. *Trans. Highl. agric. Soc. Scot.*

ON SOME MODES OF POPULATION GROWTH LEADING TO R. A. FISHER'S LOGARITHMIC SERIES DISTRIBUTION

By DAVID G. KENDALL, M.A.
Magdalen College, Oxford

1. R. A. Fisher (1943) in a co-operative study written with A. S. Corbet and C. B. Williams has developed a mathematical theory which describes with some success the relative numbers of animals of different species obtained when sampling at random from a heterogeneous population. This problem was first considered in relation to (i) Corbet's work on the distribution of butterflies in the Malay Peninsula, and (ii) the numbers of moths of different species caught in a light-trap over a given period of time (Williams's data). Fisher began by assuming that for a particular species the number of individuals caught in time t would be distributed as a Poisson variable of expectation ωt, where ω may be called the *intrinsic abundance* of the species. He suggested that ω might be distributed in the Eulerian (or χ^2) form

$$\frac{1}{\Gamma(k)}\left(\frac{k}{\Omega}\right)^k e^{-k\omega/\Omega}\omega^{k-1}\,d\omega \quad (0<\omega<\infty), \tag{1}$$

where Ω is the mean value of ω and k is a constant parameter, and showed that the actual number caught would then follow a negative binomial distribution with index k*. In fitting such a distribution to Corbet's data he obtained very small values of k, and this suggested that it might be worth while examining what would happen if Ω and k were allowed to tend to zero in a constant ratio. In this way Fisher found that if a species were known to have been caught, it would be represented in the catch by exactly n individuals with a probability

$$\frac{x^n}{ny} \quad (n = 1, 2, 3, \ldots), \tag{2}$$

where† $$y = \ln\frac{1}{1-x}, \quad x = \frac{at}{1+at} \quad \text{and} \quad a = \lim \Omega/k.$$

The success of this 'logarithmic series distribution' in graduating the entomological data of Corbet, Williams and others implies that in the populations concerned the distribution of intrinsic abundance must be (for ω not too small) effectively of the form

$$Ae^{-\omega/a}\,d\omega/\omega \quad (A = \text{constant}). \tag{3}$$

(This cannot of course be true for *all* ω, for then the integral of total probability would not converge.) It will be noticed that the distribution (3) of intrinsic abundance is itself the continuous analogue of the logarithmic series. The success of (2) in describing the relative numbers of individuals caught is thus a challenge to biologists to provide a theoretical interpretation for (3).

In this connexion it is worth noting that if one is concerned with a population containing only a finite number (Z, say) of species, then the continuous distribution (3) can be replaced by a logarithmic series, and results similar to those of Fisher follow as before. Thus, suppose that the actual number ν of individuals by which a particular species is represented *in the whole population* is distributed in the discrete form

$$\frac{X^\nu}{\nu Y} \quad (\nu = 1, 2, 3, \ldots), \tag{4}$$

* This step in the argument is of course equivalent to that taken by M. Greenwood & G. U. Yule (1920) in another context.

† I write $\ln z$ for the natural logarithm of z.

where $Y = -\ln(1-X)$, and let $p = 1-e^{-\gamma t}$ be the chance that an individual will be caught in an exposure of duration t. Then the chance that the species will have $n = 0, 1, 2, 3, \ldots$ representatives *in the sample* is given by*

$$P_0 = 1-\frac{y}{Y} \quad \text{and} \quad P_n = (1-P_0)\frac{x^n}{ny} \quad (n>0), \tag{5}$$

i.e. a logarithmic distribution with a zero term added. Here

$$x = 1-e^{-y} \quad \text{and} \quad e^y - 1 = (e^Y-1)(1-e^{-\gamma t}), \tag{6}$$

and so $y \leqslant Y$ for all t; for very small exposures p will be small and then

$$y \simeq \gamma t(e^Y - 1),$$

while for very long exposures p will be nearly equal to unity and then

$$y \simeq Y.$$

The expected values of S and N (the number of species and the number of individuals in the catch) are

$$\bar{S} = Z(1-P_0) = yZ/Y \quad \text{and} \quad \bar{N} = (e^y - 1)\,Z/Y.$$

Thus as $t \to \infty$, $\bar{S} \to Z$ and $\bar{N} \to (e^Y - 1)\,Z/Y$, while for *all* values of t

$$\bar{S} = \alpha \ln(1+\bar{N}/\alpha), \tag{7}$$

where $\alpha = Z/Y$ is a constant independent of the time of exposure and corresponding to Fisher's 'index of diversity'. Formula (7) is, in fact, identical with the well-known result due to Fisher (1943), although the derivation given here proceeds from somewhat different assumptions.

Williams (1944 a, b) has shown that the logarithmic series (2) can also be applied to a great variety of other biological problems, in which the integer n is variously the number of species per genus, the number of genera per subfamily, the number of parasites per host, and even the number of research papers per biologist (published in a particular year). It is hard to believe that a single mechanism will be found to explain the relevance of the logarithmic series to *all* these problems, and it seems therefore well worth while to record any theoretical models which may be found to lead mathematically to this distribution. In the remainder of this paper I shall describe a number of discontinuous Markoff processes which lead to distributions of negative binomial and logarithmic series form, in the hope that some of these may be found to be of biological significance.

2. The stochastic processes to be considered here will for convenience be described in relation to the growth of a hypothetical population of organisms, whose numbers fluctuate with the incidence of mortality, reproduction (by binary fission) and immigration from the outside world. Let n be the size of the population at time t; then $n(t)$ is a random function which develops in the following manner:

(i) If $n > 0$, the only possible transitions in an element of time dt are from n to $n-1$, n or $n+1$, and the transition probabilities are

$$n \to \begin{cases} n+1 & \text{with probability} \quad (n\beta+\kappa)\,dt, \\ n & \text{with probability} \quad 1-(n\mu+n\beta+\kappa)\,dt, \\ n-1 & \text{with probability} \quad n\mu\,dt. \end{cases}$$

* It appears that this sampling property of the logarithmic series distribution (which is easily proved with the aid of the generating function) has already been noticed by C. B. Williams (1947) and M. H. Quenouille.

(ii) (This is actually included in (i), but an explicit statement is desirable.) If $n = 0$, the only possible transitions in time dt are from 0 to 0 or 1, the transition probabilities being

$$0 \to \begin{cases} 1 \text{ with probability } & \kappa\, dt, \\ 0 \text{ with probability } & 1-\kappa\, dt. \end{cases}$$

Let $P_n(t)$ be the probability that at time t the population size is n; then it is possible to set up an infinite system of differential-difference equations which together with the distribution $\{P_n(t_0)\}$ at some initial time $t = t_0$ determine the $\{P_n(t)\}$ at all subsequent times, and so govern the mode of growth of the population. Two alternative sets of initial conditions will be considered here:

(A) $$P_0(-T) = 1 \quad \text{and} \quad P_n(-T) = 0 \quad (n > 0).$$

This implies that at time $t = -T$ the population size was zero.

(B) $$P_0(-T) = 0, \quad P_1(-T) = 1 \quad \text{and} \quad P_n(-T) = 0 \quad (n > 1).$$

This implies that the population commenced with one individual at time $t = -T$.

Next it is necessary to give a biological interpretation of the effects associated with the constants β, μ and κ.

The first of these, β, represents the reproductive power of the individuals composing the population, the effects of sex and age being ignored. Thus it is supposed that if attention is focused on any one individual at time t, it will be found to undergo binary fission at a time $t+\tau$, where τ has the probability distribution

$$e^{-\beta\tau}\beta\, d\tau \quad (0 < \tau < \infty). \tag{8}$$

An important consequence of the assumption (8) is that the time to the next subdivision, for any individual, is statistically independent of its past history, and in particular it is independent of the length of time since that individual was itself formed by the fission of its parent. At first sight it might appear that a bacterial colony would provide a good example of such a population growing by binary fission, but it must be remembered that the generation times of bacteria, while liable to considerable random variation, have a frequency distribution* very different from (8) and possessing a pronounced non-zero mode.

The n individuals present at any time are assumed to reproduce themselves independently of one another, and at the same constant mean rate. At each subdivision the parent can be thought of either as being replaced by its two offspring, or as only adding one new member to the colony and remaining a member itself; a transition $n \to n+1$ then takes place.

In a similar way the constant μ represents the loss to the colony due to 'mortality'. It is assumed that an individual does not lose its power to reproduce unless it 'dies', and that it then ceases to be regarded as a member of the colony, so that a transition $n \to n-1$ takes place. Such a transition could, however, also mean the removal (by any means) of an individual from the region considered; these two sources of loss are mathematically indistinguishable and will therefore be covered by the same symbol μ. Thus if an individual is observed at time t, it will disappear from the population at a time $t+\tau$, where τ has the distribution

$$e^{-\mu\tau}\mu\, d\tau \quad (0 < \tau < \infty).$$

* See, for example, Kelly & Rahn (1932) and Hinshelwood (1946).

The β- and μ-effects, described separately, are to be thought of as acting simultaneously and independently one of the other. Thus, when the β- and μ-effects are acting together, the chance that an individual remains inactive for a time τ and then subdivides during the subsequent time interval $d\tau$ is

$$e^{-(\beta+\mu)\tau}\beta\,d\tau.$$

Integration from $\tau = 0$ to $\tau = \infty$ then gives the chance that the individual will subdivide before the μ-effect has removed it from the colony; this is

$$\int_0^\infty e^{-(\beta+\mu)\tau}\beta\,d\tau = \frac{\beta}{\beta+\mu}.$$

Finally the κ-effect is one of 'immigration from outside', i.e. it is supposed that from time to time individuals not initially members of the colony may join it and proceed to behave exactly like the other members. If, from a given time instant t, the next such 'immigration' occurs at time $t+\tau$, the distribution of τ is assumed to be

$$e^{-\kappa\tau}\kappa\,d\tau \quad (0<\tau<\infty).$$

The structure of the model will now be clear. It only remains to point out that the probabilities of a positive unit increment from the β-effect or a negative unit increment from the μ-effect in an element of time dt will each be proportional to n, the existing population size, while the chance of a unit positive increment from the κ-effect is the same for all values of n.

3. The differential-difference equations of the process can now be written down. They are

$$\frac{d}{dt}P_n(t) = (n+1)\mu P_{n+1}(t) - \{n(\beta+\mu)+\kappa\}P_n(t) + \{(n-1)\beta+\kappa\}P_{n-1}(t), \tag{9}$$

if $n \geqslant 1$, and

$$\frac{d}{dt}P_0(t) = \mu P_1(t) - \kappa P_0(t). \tag{10}$$

It is convenient to *define* $P_n(t)$ as being identically equal to zero when $n<0$; equation (10) can then be included within the general form (9). I owe to Dr M. S. Bartlett the remark that systems of equations of this type can most conveniently be solved with the aid of the generating function

$$\phi(z,t) \equiv \sum_{n=-\infty}^{\infty} z^n P_n(t). \tag{11}$$

It will be seen from (9) that $\phi(z,t)$ must satisfy the partial differential equation

$$\frac{\partial\phi}{\partial t} = \{\mu - (\beta+\mu)z + \beta z^2\}\frac{\partial\phi}{\partial z} + \kappa(z-1)\phi, \tag{12}$$

which together with one of the boundary conditions,

(A) $$\phi(z,-T) = 1,$$

or

(B) $$\phi(z,-T) = z;$$

and the requirement that the expansion of ϕ must contain no terms in $1/z$, $1/z^2$, ..., is sufficient completely to determine the process.

The differential equation (12) is of the standard Lagrangian form, the auxiliary equations being

$$-dt = \frac{dz}{(\beta z-\mu)(z-1)} = -\frac{d\phi}{\kappa(z-1)\phi}. \tag{13}$$

'First integrals' are $(\beta-\mu)t+\ln(z-1)-\ln(\beta z-\mu) = \text{constant}$,

and $\kappa\ln(\beta z-\mu)+\beta\ln\phi = \text{constant}$,

if $\kappa>0$ and $\beta\neq\mu$, and so the general integral of (12) is then

$$\phi(z,t) = (\mu-\beta z)^{-\kappa/\beta}\Phi\left\{\frac{\mu-\beta z}{1-z}e^{-(\beta-\mu)t}\right\}, \tag{14}$$

where Φ is an arbitrary function to be determined from the boundary conditions.

With boundary condition (A) it will be found that

$$\phi(z,0) = \left(\frac{\beta-\mu}{\beta\Lambda-\mu}\right)^{\kappa/\beta}\left\{1-\frac{\beta(\Lambda-1)}{\beta\Lambda-\mu}z\right\}^{-\kappa/\beta}, \tag{15}$$

where Λ has been written for $e^{(\beta-\mu)T}$ (this is equal to the *expected* factor by which the population will be multiplied in a time interval T, when it is growing in the absence of immigration). Similarly with boundary condition (B) one obtains

$$\phi(z,0) = \frac{(\beta-\mu)^{\kappa/\beta}}{(\beta\Lambda-\mu)^{1+\kappa/\beta}}\{\mu(\Lambda-1)-(\mu\Lambda-\beta)z\}\left\{1-\frac{\beta(\Lambda-1)}{\beta\Lambda-\mu}z\right\}^{-1-\kappa/\beta}, \tag{16}$$

where Λ has the same meaning as before.

When $\kappa = 0$, the solutions are of a slightly different form. The general integral of (12) is then

$$\phi(z,t) = \Phi\left\{\frac{\mu-\beta z}{1-z}e^{-(\beta-\mu)t}\right\}. \tag{17}$$

Condition (A), as might be expected, gives $\phi(z,t)\equiv 1$; this merely asserts that a zero population will remain zero if there is no immigration. Condition (B), however, gives

$$\phi(z,0) = \frac{\mu(\Lambda-1)-(\mu\Lambda-\beta)z}{\beta\Lambda-\mu}\left\{1-\frac{\beta(\Lambda-1)}{\beta\Lambda-\mu}z\right\}^{-1}. \tag{18}$$

It will be noticed that in every case the solution is a regular function of z near $z = 0$, so that in the Laurent expansion the coefficients of the negative powers will all vanish, as required.

4. It is now a simple matter to interpret these solutions. There are three cases of special interest.

(i) Consider first a population growing from zero, so that (A) is the appropriate boundary condition, the population being established in the first instance by immigration from outside. The κ-effect is of course acting all the time, even after the colony has started growing, so that in general there will at any time be present a number of independent families, each descended from a different immigrant ancestor. According to (15) the population size n, after the process has been developing for a time T, will be distributed as a negative binomial variate with index κ/β and mean value

$$E(n) = \frac{\kappa(\Lambda-1)}{\beta-\mu}. \tag{19}$$

Thus if $\beta>\mu$, the expected size of the population will for large T grow geometrically at an exponential rate $(\beta-\mu)$.

If, however, $\beta<\mu$, so that the force of mortality more than compensates for the force of reproduction, one will have

$$\lim_{T\to\infty} E(n) = \frac{\kappa}{\mu-\beta}. \tag{20}$$

This is the mean of the stable distribution of population size which can just be maintained by the immigration rate κ. If κ were equal to zero the population would almost certainly die out in a finite time.

(ii) Now suppose that κ *is very small*, though still just greater than zero; to be more precise suppose that the ratio κ/β is negligible, so that while immigration is sufficient to start off the process, and to restart it if ever the population is wiped out by an excess of the μ-effect, it is negligible when compared with the contributions from the β-effect while the colony is actually growing. Then, exactly as in Fisher's analysis referred to in § 1, it will be seen that the size distribution of such a colony observed at any time will be a logarithmic series: in fact the distribution of n, *given that* $n>0$, is*

$$\frac{x^n}{ny} \quad (n = 1, 2, 3, \ldots),$$

where as always $y = -\ln(1-x)$, and

$$x = \frac{\beta(\Lambda-1)}{\beta\Lambda-\mu}. \tag{21}$$

If T is very large, then

$$x \simeq 1-\left(1-\frac{\mu}{\beta}\right)\frac{1}{\Lambda} \quad \text{when} \quad \beta>\mu, \tag{22}$$

and

$$x \simeq \beta/\mu \quad \text{when} \quad \beta<\mu. \tag{23}$$

The second case is the more interesting, for one can then let T tend to infinity and so obtain the stable distribution of population size when $\beta<\mu$. Of course in the limit when $\kappa = 0$ it is 'almost certain' that $n = 0$. If, however, a colony *does* exist (i.e. if $n>0$), then it is almost certainly homogeneous (descended from a single immigrant ancestor), and its size n will be distributed in a logarithmic series.

(iii) The case $\beta = \mu$, when reproduction and mortality just balance, requires a separate discussion. The equations auxiliary to (12) are now

$$-dt = \frac{dz}{\beta(z-1)^2} = -\frac{d\phi}{\kappa(z-1)\phi},$$

and the general integral is

$$\phi(z,t) = (1-z)^{-\kappa/\beta}\Phi\left\{\beta t+\frac{1}{1-z}\right\}.$$

Condition (A) then gives

$$\phi(z,0) = (1+\beta T)^{-\kappa/\beta}\left\{1-\frac{\beta T}{1+\beta T}z\right\}^{-\kappa/\beta}, \tag{24}$$

so that here again n is distributed as a negative binomial variate with index κ/β, but the mean value is now

$$E(n) = \kappa T, \tag{25}$$

so that the expected size of the population is linearly proportional to the time of exposure to immigration. If κ/β is very small the limiting conditional distribution of n (given that $n>0$) is once again the Fisher series

$$\frac{x^n}{ny} \quad (n = 1, 2, 3, \ldots),$$

where now

$$x = \frac{\beta T}{1+\beta T}. \tag{26}$$

* This conditional distribution is of course obtained by taking the ratio of the general term to the sum of all the terms but the first, in the negative binomial series, and then letting $\kappa/\beta \to 0$.

5. It will perhaps have been noticed that in the analysis of the last section the immigration rate κ (when κ/β is small) merely has the effect of ensuring that an observed population is almost certainly descended from a single immigrant ancestor who entered the region considered at some instant during the time interval of length T preceding the moment of observation. I think this gives us a clue to the true 'explanation' of the occurrence of the logarithmic series in the solution to the problem just considered. Before exploring the matter further it will be found helpful to examine the growth of a single family by setting $\kappa = 0$ and starting from unit population at time $t = -T$, thus employing boundary condition (B).

From the generating function (18) it then follows that the distribution of the population size n at the time of observation ($t = 0$) is*

$$P_0 = \frac{\mu(\Lambda-1)}{\beta\Lambda-\mu} \quad \text{and} \quad P_n = (1-P_0)(1-u)u^{n-1} \quad (n>0), \tag{27}$$

where

$$u = \frac{\beta(\Lambda-1)}{\beta\Lambda-\mu}.$$

The distribution is thus a geometric series with a modified zero term, the mean population size being

$$E(n) = \Lambda,$$

so that the *expected* population grows geometrically at an exponential rate $\beta-\mu$ (which will, of course, be negative if $\beta<\mu$).

It is of interest to evaluate the variance of the population size. This proves to be

$$\operatorname{Var}(n) = E(n-\Lambda)^2 = \frac{\beta+\mu}{\beta-\mu}\Lambda(\Lambda-1), \tag{28}$$

and the coefficient of variation of the population size is thus

$$\sqrt{\left(\frac{\beta+\mu}{\beta-\mu}\right)}\sqrt{\left(1-\frac{1}{\Lambda}\right)} \quad \text{if} \quad \beta>\mu$$

and

$$\sqrt{\left(\frac{\mu+\beta}{\mu-\beta}\right)}\sqrt{\left(\frac{1}{\Lambda}-1\right)} \quad \text{if} \quad \beta<\mu.$$

For large T these expressions become

$$\text{C. of V.}(n) \simeq \sqrt{\left(\frac{\beta+\mu}{\beta-\mu}\right)} \quad \text{if} \quad \beta>\mu, \tag{29}$$

and

$$\text{C. of V.}(n) \simeq \sqrt{\left\{\frac{\mu+\beta}{\mu-\beta}\frac{1}{E(n)}\right\}} \quad \text{if} \quad \beta<\mu; \tag{30}$$

in the second case it will be recalled that $E(n)\to 0$ as $T\to\infty$.

The distribution of the population size thus behaves rather differently in the two cases of an exponentially growing and an exponentially decreasing population; this agrees with a conclusion reached by M. S. Bartlett (1937). He considered the similar problem when there is no spreading of generations and the generation-time is rigorously constant. As he points out, there is a connexion with Fisher's theory of the extinction of rare characters and, one might add, with the work of Francis Galton, H. W. Watson, and A. J. Lotka on the extinction of surnames.†

* (*Added in proof.*) Dr Bartlett has pointed out to me that the result (27) is stated by N. Arley & V. Borchsenius in *Acta Math.* (1945), **76**, 298–9. It is attributed by them to Dr C. Palm.

† See Fisher (1930), Galton (1889) and Lotka (1931). The problem of the distribution of surnames, and its variation in time, seems not yet to have received all the attention it deserves. The surname is a 'rare character' whose extinction can very readily be observed; normal social conditions ensure that it is inherited as if it were controlled by a gene totally sex-linked in Y. Reference may be made to the work of R. A. Fisher & Janet Vaughan (1939), and J. A. Fraser Roberts (1941–2), who have considered the relation between surnames and blood-groups. [See also reference to Tocher & Gray, p. 1 above. ED.]

When $\beta = \mu$, so that the forces of reproduction and mortality just balance, the above solution must be modified. The appropriate results are most easily obtained by letting β tend to μ in the several formulae. In this way one finds

$$P_0 = \frac{\beta T}{1+\beta T} \quad \text{and} \quad P_n = (1-P_0)(1-u)u^{n-1} \quad (n>0), \tag{31}$$

where
$$u = \frac{\beta T}{1+\beta T},$$

and so
$$E(n) = 1, \quad \mathrm{Var}(n) = 2\beta T \quad \text{and} \quad \text{C. of V.}(n) = \sqrt{(2\beta T)}. \tag{32}$$

It is of interest to note that if β and μ instead of being constants are each proportional to the *same* function of the time,* the above theory still holds, provided that T is everywhere replaced by

$$\int_{-T}^{0} \psi(t)\,dt,$$

where $\beta = \beta_0 \psi(t)$ and $\mu = \mu_0 \psi(t)$.

6. Consider now the size distribution of a colony developing in the absence of immigration and known to have originated from a single individual whose arrival in the region concerned occurred during the preceding T time-units. If the time of arrival of the common ancestor is a random variable uniformly distributed from $t = -T$ to $t = 0$, it will follow from (27) that

$$P_n = \frac{1}{\beta T}\int_0^T \left\{\frac{\beta(\lambda-1)}{\beta\lambda-\mu}\right\}^{n-1} \frac{\beta\lambda(\beta-\mu)^2}{(\beta\lambda-\mu)^2}\,d\tau \quad (n \geqslant 1),$$

where $\lambda = e^{(\beta-\mu)\tau}$. Now since
$$\frac{d}{d\tau}\left\{\frac{\beta(\lambda-1)}{\beta\lambda-\mu}\right\} = \frac{\beta\lambda(\beta-\mu)^2}{(\beta\lambda-\mu)^2},$$

this can be written
$$P_n = \frac{1}{\beta T}\int_0^U u^{n-1}\,du = \frac{U^n}{n\beta T},$$

while
$$P_0 = 1 + \frac{1}{\beta T}\ln(1-U),$$

U being the value of
$$u = \frac{\beta(\lambda-1)}{\beta\lambda-\mu}$$

when $\tau = T$. Thus the distribution of the population size at the time of observation ($t = 0$) is

$$P_0 = 1 - \frac{1}{\beta T}\ln\left(\frac{1}{1-x}\right) \quad \text{and} \quad P_n = (1-P_0)\frac{x^n}{ny} \quad (n>0), \tag{33}$$

where
$$x = \frac{\beta(\Lambda-1)}{\beta\Lambda-\mu}, \tag{34}$$

y and Λ being defined as before. For large T,

$$x \simeq 1 - \left(1-\frac{\mu}{\beta}\right)\frac{1}{\Lambda} \quad \text{if} \quad \beta>\mu, \tag{35}$$

and
$$x \simeq \beta/\mu \quad \text{if} \quad \beta<\mu. \tag{36}$$

In the latter case it is permissible to let T tend to infinity and so obtain the stable (logarithmic series) distribution for n (*given that* $n>0$). When $\beta = \mu$ the (x, T) relation is $x = \beta T/(1+\beta T)$.

It will now be clear why, in § 4 (ii), the boundary condition (A) together with the hypothesis $n>0$ led to a logarithmic series distribution for n as $\kappa/\beta \to 0$. For in these circumstances it

* A discussion of the similar problem when β and μ are *any* (not necessarily the same) functions of the time will be given in my paper (1948).

would be almost certain that an observed population was wholly descended from a single immigrant ancestor who arrived in the field of observation at an unknown time instant uniformly distributed between $t = -T$ and $t = 0$.*

7. The classical theory of population growth (see, for example, A. J. Lotka (1945) for a general review and extensive references) is largely based on a *deterministic* description of the phenomena, which leads to differential and integral equations for the expectation values of the random variables concerned (the total population size, and the numbers in the several age groups). Apart from the work on 'extinction' already mentioned, the first *stochastic* treatment of the general problems of population growth seems to be that of W. Feller in an important paper† which is unfortunately not generally accessible in this country. Feller's work has been further developed by N. Arley (1943),‡ who showed that discontinuous Markoff processes of a similar type are equally relevant in the theory of the 'cascade showers' initiated by cosmic ray particles. In particular he makes use of the simple birth-and-death process discussed here in § 5, and quotes Feller's formulae for the mean and variance of n as functions of the time t. Arley gives an elegant method for determining an expansion for $P_n(t)$ in powers of $(t-t_0)$, and observes that the calculation of high order coefficients becomes very cumbersome. One can now see, on examining the complete solution contained in (27), that this is very reasonable, for $P_n(t)$ is quite a complicated function of the time t, although a simple one of the reduced variable u.

The results of the present paper would not have been obtained without the generating-function technique for transforming the differential-difference equations into a partial differential equation of simple type. This device was suggested to me by Dr M. S. Bartlett in the summer of 1946, and he has also applied his method to a number of Markoff processes of interest in biology, one of which is the birth-and-death process of § 5. An account of this work is now available (Bartlett, 1947). In another place I intend to give a discussion of the most general birth-and-death process of this type, in which the birth- and death-rates β and μ can be any functions of the time t (Kendall, 1948); this development also makes use of the generating-function procedure.

It is a pleasure to acknowledge my indebtedness to Dr Bartlett and to the many other friends whose comments have helped to clarify my ideas on this subject. In particular, I should like to thank Mr D. J. Finney and members of his Seminar in Biological Statistics in Oxford, and Mr F. J. Anscombe, Dr P. Jones and Dr C. B. Williams of the Rothamsted Experimental Station.

* The condition $n > 0$ is to be introduced *after* (33) has been established. If it were imposed from the start, the *a posteriori* distribution of the time of arrival of the immigrant ancestor would no longer be a uniform one, but the appropriate modification of the argument would lead to the same final result.

† Feller (1939). I am greatly indebted to Prof. Feller for the gift of a reprint of this paper. In private correspondence he tells me that he also, in unpublished lectures, has solved the equations governing the birth-and-death process discussed here in § 5.

‡ Arley's monograph is also of great value in presenting a useful account (including several new developments) of the Kolmogoroff-Feller theory of Markoff processes, especially those of the discontinuous type used here. Another excellent account has been given recently by O. Lundberg (1940). For a general introduction to this subject, reference may be made to my review article (1947).

REFERENCES

ARLEY, N. (1943). *On the Theory of Stochastic Processes and their Application to the Theory of Cosmic Radiation.* Copenhagen: G. E. C. Gads.

BARTLETT, M. S. (1937). Deviations from expected frequencies in the theory of inbreeding. *J. Genet.* **35**, 83–7.

BARTLETT, M. S. (1947). *Stochastic Processes.* (Notes of a course given at the University of North Carolina in the Fall Quarter, 1946. It is understood that copies of these notes are available on request.)

FELLER, W. (1939). Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in Wahrscheinlichkeitstheoretischer Behandlung. *Acta Biotheoretica*, **5**, 11–40. This seems not to be generally available in Great Britain, but there is an abstract in *Math. Rev.* **1**, 22.

FISHER, R. A. (1930). *The Genetical Theory of Natural Selection*, pp. 73–83. Oxford: Clarendon Press.

FISHER, R. A. & VAUGHAN, JANET (1939). Surnames and blood-groups. *Nature, Lond.*, **144**, 1047.

FISHER, R. A., CORBET, A. S. & WILLIAMS, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12**, 42–58.

GALTON, FRANCIS (1889). *Natural Inheritance*, Appendix F (which includes the contribution by H. W. WATSON). London: Macmillan.

GREENWOOD, M. & YULE, G. U. (1920). *J. R. Statist. Soc.* **83**, 255–79.

HINSHELWOOD, C. N. (1946). *The Chemical Kinetics of the Bacterial Cell*, Chapter x (see especially Fig. 64). Oxford: University Press.

KELLY, C. D. & RAHN, O. (1932). The growth rate of individual bacterial cells. *J. Bacteriol.* **23**, 147–53.

KENDALL, D. G. (1947). A review of some recent work on discontinuous Markoff processes with applications to biology, physics, and actuarial science. *J.R. Statist. Soc.* **110**, 130–7.

KENDALL, D. G. (1948). On the generalized 'birth-and-death' process. *Ann. Math. Statist.* (in the Press).

LOTKA, A. J. (1931). The extinction of families. *J. Wash. Acad. Sci.* **21**, 377–80 and 453–9.

LOTKA, A. J. (1945). Population analysis as a chapter in the mathematical theory of evolution. Included in *Essays on Growth and Form presented to D'Arcy Wentworth Thompson* (edited by Le Gros Clark, W. E. & Medawar, P. B.). Oxford: Clarendon Press.

LUNDBERG, O. (1940). *On Random Processes and their Application to Sickness and Accident Statistics.* Uppsala: Almqvist and Wiksells.

ROBERTS, J. A. FRASER (1941–2). Blood-group frequencies in north Wales. *Ann. Eugen., Lond.*, **11**, 260–71. (Includes references to earlier work on the distribution of surnames.)

WILLIAMS, C. B. (1944*a*). Some applications of the logarithmic series and the index of diversity to ecological problems. *J. Ecol.* **32**, 1–44.

WILLIAMS, C. B. (1944*b*). The numbers of publications written by biologists. *Ann. Eugen., Lond.*, **12**, 143–6.

WILLIAMS, C. B. (1947). The logarithmic series and its application to biological problems. *J. Ecol.* (in the Press).

THE STUDENTIZED FORM OF THE EXTREME MEAN SQUARE TEST IN THE ANALYSIS OF VARIANCE†

BY K. R. NAIR

CONTENTS

PART I. EXPANSION OF THE STUDENTIZED INTEGRAL: EXTENSION OF H. O. HARTLEY'S RESULTS

1. *Introduction*

Let $x_1, \ldots, x_n$ denote a sample of n observations drawn from a normal population with mean μ and standard deviation σ. Then the difference between sample and population mean, $\bar{x}-\mu$, is normally distributed with zero mean and standard deviation $\sigma/\sqrt{n}$. If we estimate the standard deviation of the parent, σ, by

$$s = \sqrt{[\Sigma(x-\bar{x})^2/(n-1)]},$$

then 'Student' (1908) gave the distribution of

$$t = (\bar{x}-\mu)\sqrt{n}/s.$$

As was to be expected, this distribution was independent of σ. The knowledge of the distribution of t made it possible to draw inferences, with the help of evidence entirely supplied by the sample, about the location-parameter μ of a normal population, without making any assumptions about the generally unknown scale-parameter, σ.

Neyman & Pearson (1928) extended this notion to an analogous problem connected with the rectangular and exponential populations.

Fisher (1924) obtained the distribution of s'/s where s' and s are two independent estimates of σ, calculated by the root-mean-square method. This distribution is also independent of σ. Fisher also showed that 'Student's' t-distribution could be extended to a wide range of sampling problems. Sukhatme (1937) has shown that analogues to t and s'/s of 'normal theory' can be developed for an exponential population, again eliminating the unknown scale parameter.

This notion of eliminating unknown scale parameters from the distribution laws of statistics has come to be known as 'studentization'.

† Part of a thesis approved for the degree of Ph.D. of the University of London.

If instead of s' and s which give unbiased estimates of σ, we use other types of statistics which are 'proportional' to σ,* such as the range, mean deviation, etc., it is clear that the distribution of the ratio of two such estimates will be independent of σ, although the analytic expression for this distribution may be difficult to obtain.

Hartley (1938, 1944) has shown that if instead of s'/s we have a ratio w/s, where (i) w is a general statistic 'proportional' to σ calculated from a sample of n observations $x_1, \ldots, x_n$ drawn from a normal population and (ii) s is independently distributed with ν degrees of freedom, then the distribution of w/s can be derived without much difficulty if the distribution of w/σ is known. His solution is described in some detail in the next section and its application to special problems is considered in the succeeding sections.

The probability integral of w/s may be called the *studentized integral* of w. When $\nu \to \infty$ the studentized integral becomes identical with the probability integral of w/σ, which may in turn be called the ∞-*integral* of w/s.

2. *Expansion of the studentized integral in powers of* ν^{-1}

Let $f_\nu(w/s)$ denote the distribution function of w/s. When $\nu \to \infty$ this gives the distribution function of w/σ, which we shall denote by $f(w/\sigma)$.

Let the probability integral of w/s be

$$P_\nu(Q) = \int_0^Q f_\nu(w/s)\, d(w/s).$$

When $\nu \to \infty$, $P_\nu(Q)$ gives the probability integral of w/σ which we may denote as $P(Q)$.

Assuming σ to be unity, it follows (see, for example, Hartley) that

$$P_\nu(Q) = 2\Gamma(\tfrac{1}{2}\nu)^{-1}(\tfrac{1}{2}\nu)^{\frac{1}{2}\nu}\int_0^\infty s^{\nu-1}e^{-\frac{1}{2}\nu s^2}P(Qs)\,ds. \tag{1}$$

Hartley made the significant discovery that a recurrence formula connecting $P_\nu(Q)$ and $P_{\nu-2}(Q)$ could be built up and that, thereby, $P_\nu(Q)$ can be obtained as the solution of a partial differential equation. He solves this equation by iteration, and at the third stage obtains the first three terms in an expansion for $P_\nu(Q)$ in powers of ν^{-1}, namely,

$$P_\nu(Q) = P(Q) + \frac{1}{4\nu}(Q^2P'' - QP') + \frac{1}{6(4\nu)^2}(3Q^4P^{\text{iv}} - 2Q^3P''' - 3Q^2P'' + 3QP'), \tag{2}$$

where the derivatives of P are taken at the argument Q.

For ν of the order 20 or larger this formula will in most cases give quite sufficient accuracy. Sometimes when ν is small, however, we may need at least the next two terms of the expansion. The term in ν^{-3} immediately follows from equation (24) of Hartley's (1944) paper and is

$$\frac{e^{-3x}}{12}\left(\tfrac{1}{32}\phi^{\text{vi}} - \tfrac{7}{16}\phi^{\text{v}} + \tfrac{17}{8}\phi^{\text{iv}} - \tfrac{17}{4}\phi''' + 3\phi''\right), \tag{3†}$$

where $\phi(\lambda) = P(e^\lambda) = P(Q)$ and $x = \log \nu$.

The term in ν^{-4} will come from the fifth stage of the iteration which involves some lengthy algebra. In terms of x, ϕ and λ this becomes

$$\frac{e^{-4x}}{360(4^4)}(15\phi^{\text{viii}} - 360\phi^{\text{vii}} + 3320\phi^{\text{vi}} - 14784\phi^{\text{v}} + 32240\phi^{\text{iv}} - 28800\phi''' + 1280\phi'' + 6144\phi'). \tag{4}$$

* A statistic w is 'proportional' to σ if it transforms into w/σ when the observations x_i are measured in units of σ.

† The coefficient of ϕ''' is misprinted $-\frac{11}{6}$ for $-\frac{17}{4}$ in Hartley's paper.

The derivatives of $\phi(\lambda)$ can be expressed in terms of those of $P(Q)$.

$$\left.\begin{aligned}
\phi(\lambda) &= P(Q),\\
\phi' &= QP',\\
\phi'' &= Q^2P''+QP',\\
\phi''' &= Q^3P'''+3Q^2P''+QP',\\
\phi^{\mathrm{iv}} &= Q^4P^{\mathrm{iv}}+6Q^3P'''+7Q^2P''+QP',\\
\phi^{\mathrm{v}} &= Q^5P^{\mathrm{v}}+10Q^4P^{\mathrm{iv}}+25Q^3P'''+15Q^2P''+QP',\\
\phi^{\mathrm{vi}} &= Q^6P^{\mathrm{vi}}+15Q^5P^{\mathrm{v}}+65Q^4P^{\mathrm{iv}}+90Q^3P'''+31Q^2P''+QP',\\
\phi^{\mathrm{vii}} &= Q^7P^{\mathrm{vii}}+21Q^6P^{\mathrm{vi}}+140Q^5P^{\mathrm{v}}+350Q^4P^{\mathrm{iv}}+301Q^3P'''+63Q^2P''+QP',\\
\phi^{\mathrm{viii}} &= Q^8P^{\mathrm{viii}}+28Q^7P^{\mathrm{vii}}+266Q^6P^{\mathrm{vi}}+1050Q^5P^{\mathrm{v}}+1701Q^4P^{\mathrm{iv}}+966Q^3P'''\\
&\quad +127Q^2P''+QP'.
\end{aligned}\right\} \tag{5}$$

Substituting these in (3) and (4) we get, up to terms in ν^{-4},

$$P_\nu(Q) = a_0+a_1/\nu+a_2/\nu^2+a_3/\nu^3+a_4/\nu^4, \tag{6}$$

where

$$\begin{aligned}
a_0 &= P(Q),\\
a_1 &= \tfrac{1}{4}(Q^2P''-QP'),\\
a_2 &= \frac{1}{6(4^2)}(3Q^4P^{\mathrm{iv}}-2Q^3P'''-3Q^2P''+3QP'),\\
a_3 &= \frac{1}{6(4^3)}(Q^6P^{\mathrm{vi}}+Q^5P^{\mathrm{v}}-7Q^4P^{\mathrm{iv}}+12Q^3P'''-15Q^2P''+15QP'),\\
a_4 &= \frac{1}{360(4^4)}(15Q^8P^{\mathrm{viii}}+60Q^7P^{\mathrm{vii}}-250Q^6P^{\mathrm{vi}}+366Q^5P^{\mathrm{v}},\\
&\qquad -285Q^4P^{\mathrm{iv}}-30Q^3P'''+945Q^2P''-945QP').
\end{aligned} \tag{7}$$

From (7) we can derive the following results

$$\left.\begin{aligned}
\int QP'dQ &= -\frac{4}{3}\int a_1dQ = -\frac{32}{25}\int a_2dQ = -\frac{128}{105}\int a_3dQ = -\frac{2048}{1659}\int a_4dQ,\\
\int Q^2P'dQ &= -\int Qa_1dQ = -\frac{1}{2}\int Qa_2dQ = -\frac{1}{4}\int Qa_3dQ = -\frac{1}{8}\int Qa_4dQ,\\
\int Q^3P'dQ &= -\frac{4}{5}\int Q^2a_1dQ = -\frac{96}{385}\int Q^2a_2dQ = -\frac{128}{1575}\int Q^2a_3dQ = -\frac{2048}{76153}\int Q^2a_4dQ,\\
\int Q^4P'dQ &= -\frac{2}{3}\int Q^3a_1dQ = -\frac{1}{7}\int Q^3a_2dQ = -\frac{1}{30}\int Q^3a_3dQ = -\frac{1}{124}\int Q^3a_4dQ,
\end{aligned}\right\} \tag{8}$$

where the limits of integration are 0 and ∞.

These results enable us to express the moments of the studentized statistic in terms of those of the non-studentized statistic. Thus, up to terms in ν^{-4}, we have

$$\left.\begin{aligned}
\int_0^\infty P_\nu(Q)\,dQ &= \int_0^\infty P(Q)\,dQ-\left(\frac{3}{4\nu}+\frac{25}{32\nu^2}+\frac{105}{128\nu^3}+\frac{1659}{2048\nu^4}\right)\int_0^\infty QP'(Q)\,dQ,\\
\int_0^\infty QP_\nu(Q)\,dQ &= \int_0^\infty QP(Q)\,dQ-\left(\frac{1}{\nu}+\frac{2}{\nu^2}+\frac{4}{\nu^3}+\frac{8}{\nu^4}\right)\int_0^\infty Q^2P'(Q)\,dQ,\\
\int_0^\infty Q^2P_\nu(Q)\,dQ &= \int_0^\infty Q^2P(Q)\,dQ-\left(\frac{5}{4\nu}+\frac{385}{96\nu^2}+\frac{1575}{128\nu^3}+\frac{76153}{2048\nu^4}\right)\int_0^\infty Q^3P'(Q)\,dQ,\\
\int_0^\infty Q^3P_\nu(Q)\,dQ &= \int_0^\infty Q^3P(Q)\,dQ-\left(\frac{3}{2\nu}+\frac{7}{\nu^2}+\frac{30}{\nu^3}+\frac{124}{\nu^4}\right)\int_0^\infty Q^4P'(Q)\,dQ.
\end{aligned}\right\} \tag{9}$$

3. *Application to the incomplete beta-function*

By applying expansion (2) to the probability integral of the square root of a variance ratio having $2p$ and $2q$ degrees of freedom, Hartley obtained a new formula for the incomplete beta-function

$$I_x(p,q) = \frac{\Gamma(p+q)}{\Gamma(p)\,\Gamma(q)}\int_0^x x^{p-1}(1-x)^{q-1}\,dx$$

in terms of the incomplete gamma-function and its derivatives.

Using the additional terms in ν^{-3} and ν^{-4} given in (6) this formula becomes

$$I_x(p,q) = \Gamma(p)^{-1}\int_0^\omega e^{-y}y^{p-1}dy + \Gamma(p)^{-1}e^{-\omega}\omega^p\{b_1/(2q)+b_2/(2q)^2+b_3/(2q)^3+b_4/(2q)^4\}, \tag{10}$$

where $x = \omega/(\omega+q)$ and

$$\left.\begin{aligned}
b_1 &= \{(p-1)-\omega\},\\
b_2 &= \tfrac{1}{6}\{(p-1)(p-2)(3p-1)-(p-1)(9p-2)\omega+(9p-1)\omega^2-3\omega^3\},\\
b_3 &= \tfrac{1}{6}\{p(p-1)^2(p-2)(p-3)-p(p-1)(p-2)(5p-3)\omega+2p(p-1)(5p-1)\omega^2\\
&\quad -2p(5p+1)\omega^3+(5p+3)\omega^4-\omega^5\},\\
b_4 &= \tfrac{1}{360}\{(p-1)(p-2)(p-3)(p-4)(15p^3-30p^2+5p+2)\\
&\quad -(p-1)(p-2)(p-3)(105p^3-135p^2+10p+8)\omega\\
&\quad +(p-1)(p-2)(315p^3-180p^2+5p+12)\omega^2\\
&\quad -(p-1)(525p^3+75p^2+50p+8)\omega^3\\
&\quad +(525p^3+450p^2+175p+2)\omega^4\\
&\quad -5(63p^2+99p+46)\omega^5+15(7p+9)\omega^6-15\omega^7\}.
\end{aligned}\right\} \tag{11}$$

This formula strictly applies only when p and q of the incomplete beta-function are such that $2p$ and $2q$ are positive integers. It is not suitable when p is large, because of the slow convergence of $b_i/(2q)^i$. For small values of p and moderate or large values of q, however, the formula is quite useful. For large p and q, Wishart (1927) and others have developed useful methods.

To give some idea of the accuracy of the formula when only terms up to b_2 are used, Hartley considered the example $p = 1$, $\omega = \frac{50}{9}$ and various integral values of q in the range 5 to 50. The agreement between the exact and the approximate values was very good for values of $q > 20$. For lower values of q, addition of the terms in b_3 and b_4 makes the approximations closely agree with the exact values.

Table giving exact and approximate values of $I_x(p,q)$ for $p = 1$, $\omega = \frac{50}{9}$ and $x = \omega/(\omega+q)$

q	Exact	Approximation		
		(1)	(2)	(3)
5	0·976 155	0·974 627	0·975 597	0·976 788
10	0·987 945	0·987 774	0·987 896	0·987 970
15	0·991 140	0·991 093	0·991 129	0·991 144
20	0·992 571	0·992 553	0·992 568	0·992 572

In this table approximation (1) uses terms up to b_2, and approximations (2) and (3) are obtained by including terms up to b_3 and b_4 respectively. The last approximation gives values in *excess* of the exact values, which shows that the value of the omitted remainder terms for $\omega = \frac{50}{9}$ is negative, though negligible when q exceeds 10.

4. *Application to 'Student's' integral*

Fisher (1926) gave an expansion for the probability integral of 'Student's' t, in powers of ν^{-1}. He worked out the coefficients of the terms up to ν^{-5} and found that the maximum value of the fifth correction never exceeded 10^{-5} when ν was greater than 18. It will be interesting to compare the coefficients in Fisher's expansion with those obtained by applying Hartley's method.

Since t is symmetrically distributed about 0, we may consider the probability integral of $|t|$. To get the expansion of this integral by Hartley's method we have only to replace the $P(Q)$ in (6) by the '∞-integral' of $|t|$, namely,

$$P(Q) = \sqrt{\frac{2}{\pi}}\int_0^Q e^{-\frac{1}{2}t^2}\,d\,|t|. \tag{12}$$

Alternatively, since $|t|$ is the same as the square root of a variance ratio having 1 and ν degrees of freedom, putting $p = \frac{1}{2}$, $q = \frac{1}{2}\nu$ and $\omega = \frac{1}{2}t^2$ in (10), we may obtain its probability integral in the form

$$\sqrt{\frac{2}{\pi}}\int_0^{|t|} e^{-\frac{1}{2}x^2}\,dx - \sqrt{\frac{2}{\pi}}\,|t|\,e^{-\frac{1}{2}t^2}\sum_{r=1}^{4} T_r\nu^{-r}, \tag{13}$$

where

$$\left.\begin{aligned}
T_1 &= \tfrac{1}{4}(t^2+1),\\
T_2 &= \frac{1}{6(4^2)}(3t^6-7t^4-5t^2-3),\\
T_3 &= \frac{1}{6(4^3)}(t^{10}-11t^8+14t^6+6t^4-3t^2-15),\\
T_4 &= \frac{1}{360(4^4)}(15t^{14}-375t^{12}+2225t^{10}-2141t^8-939t^6-213t^4+915t^2+945).
\end{aligned}\right\} \tag{14}$$

The probability integral of t from t to ∞ immediately follows as

$$\frac{1}{\sqrt{(2\pi)}}\int_t^{\infty} e^{-\frac{1}{2}x^2}\,dx + \frac{1}{\sqrt{(2\pi)}}\,te^{-\frac{1}{2}t^2}\sum_{r=1}^{4} T_r\nu^{-r}. \tag{15}$$

The coefficients T_1, T_2, T_3 and T_4* are identical to those given by Fisher providing us with a useful confirmation of the general applicability of Hartley's method.

5. *Some further applications*

In his 1938 paper, Hartley examined the studentized probability integral of the largest and smallest among k independent estimates of a variance, each having 1 degree of freedom. The importance of this problem in the analysis of factorial experiments was first pointed out by Wishart (1938). As an expansion of the integral in powers of ν^{-1} was not then known, Hartley approached the problem by an approximate method which was satisfactory for getting the lower percentage points of the smallest variance, but not for the more important case of upper percentage points of the largest variance.

* The coefficient of t^2 in T_4 appears with a negative sign in Fisher's (1926) paper but the correct sign has been used in a later paper, viz. *Ann. Eugen.* **11**, 141–72.

Meanwhile, Finney (1941) gave an exact solution for the studentized integral of the largest and smallest of k variances, each estimated with m degrees of freedom. It is not easy to make this exact solution amenable to numerical calculation of the percentage points for odd values of m. Hartley's expansion (2) with the additional terms given in (6) could usefully be employed to get the percentage points when $m = 1$. Details of the method and tables of the 5 and 1 % points are given in Part II.

Pearson & Hartley (1943) used expansion (2) to calculate the studentized integral of the range. This is the probability integral of $(x_n - x_1)/s$, where x_1 and x_n are the two extreme observations in a sample of n observations and s is an independent estimate of σ based on ν degrees of freedom. The studentized range is of importance in quality control charts for industrial products. A closely allied statistic, namely, the studentized extreme deviate $(x_n - \bar{x})/s$ (or $(\bar{x} - x_1)/s$) has useful applications in designed experiments when we are concerned with judging whether a single outlying treatment (best or worst) really differs from the rest. The probability integral of this statistic will be considered in another paper (Nair, 1948) and tables provided for $n = 3, 4, \ldots, 9$.

There are many other examples where the expansion of the studentized integral can usefully be employed. The applications cited above are convincing proof of its potentialities.

PART II. APPLICATION TO TESTS REGARDING THE LARGEST AND SMALLEST OF SEVERAL VARIANCES

1. *Introduction*

Let $s_1^2, \ldots, s_k^2$ be independent estimates of an unknown variance σ^2 of a normal population each calculated from sums of squares having m degrees of freedom and arranged in ascending order of magnitude. Let s_0^2 be another independent estimate based on ν degrees of freedom against which each of the first k estimates is to be tested for significant differences from s_0^2. In routine analysis of variance where such tests frequently occur, it is customary to test each of the k variances against s_0^2 and separately declare whether there is a significant difference or not. Forming now for each of the independent s_i^2 the ratio $F_i = s_i^2/s_0^2$, it is obvious that the largest ratio, $F_k = s_k^2/s_0^2$, is more likely to be declared significant than any other variance ratio in the set. Indeed, for $k = 20$ it would be expected to be significant at the 5 % level of F for m and ν degrees of freedom. This source of bias can be eliminated if the probability integral of the largest variance ratio is numerically evaluated.

Although the theory discussed in this paper applies to any value of m, owing to practical difficulties, tables of the upper 5 and 1 % points of the largest variance ratio have been prepared only when $m = 1$, which is the most important case in practice. Thus, as Wishart (1938) has pointed out, it has useful application in the analysis of variance for $2 \times 2 \times 2 \times \ldots$ factorial design. It could in fact be applied to the analysis of any designed experiment where the total variance due to treatment effects is split up into components each having a single degree of freedom.

Another possible application is in the fitting of curves with orthogonal terms, e.g. orthogonal polynomials and harmonic analysis. It may well happen that while, taken as a whole there is no significant curvilinearity, the coefficient of a single high order term turns out to be very large compared to the rest. The new tables will be useful in testing whether such isolated coefficients are significantly large.

The theory developed for the largest variance ratio could easily be extended to any other *ranked* ratio. Of these the smallest, namely, $F_1 = s_1^2/s_0^2$ has been considered. Occasions to test the significance of the smallest of k variance ratios are much less frequent than that of the largest. In the former case, we are concerned with the *lower* percentage points of the probability integral, in order that a test could be made whether an observed smallest ratio is significantly small.

In field experiments, the smallest of k variances tested may become significantly small compared to the 'error' variance (s_0^2) if the $m+1$ groups of plot values from which the m degrees of freedom of that variance were obtained showed a high negative intra-class correlation. The extreme value that this correlation can have, if there are l plots in each group, is $-1/(l-1)$, so that it can be -1 only if $l = 2$. Wishart (1938) again gives an example where a test for significance of smallest variance ratio when $m = 1$ is appropriate.

A rather interesting use for the test of significance of the smallest variance ratio, when $m = 1$, can be found in certain methods of statistical control in sample surveys devised by Mahalanobis (1944). If, say, you are sampling agricultural fields in k districts and send two investigators to each district to collect half the quota of sample units into two randomly selected half samples A and B, a comparison of the A and B mean values for each district should give a clue as to whether the data have been properly collected. If one of the A-B differences is very large and the variance ratio significant it is usual to conclude that the two investigators concerned did not carry out the instructions in the same manner or that some other personal bias had crept in. On the other hand if one of the k differences between (A, B) pairs is surprisingly small and turns out to give a significant smallest variance ratio, it may perhaps arouse a suspicion that the two investigators consulted each other and dishonestly made the means of their data agree. This is the negative side of the argument in favour of a test of significance of the smallest variance ratio. On the positive side, it may often turn out that the smallest variance ratio is *not* significantly small, avoiding awkward aspersions on the reliability of the investigators.

2. *Probability integral of the largest variance ratio*

Let ${}_\nu P_k(Q)$ denote the probability of s_k/s_0 being $\leqslant Q$. This is the same as the probability of $F_k = s_k^2/s_0^2$ being $\leqslant Q^2$. When $\nu \to \infty$, this probability will be denoted by $P_k(Q)$.

Finney (1941) showed that

$$ {}_\nu P_k(Q) = M^k(1+\lambda)^{-\frac{1}{2}\nu}, \tag{1}$$

where

$$M = \Gamma(\tfrac{1}{2}m)^{-1} \int_0^{-\frac{m}{\nu}Q^2\frac{\partial}{\partial\lambda}} u^{\frac{1}{2}m-1} e^{-u}\, du, \tag{2*}$$

and λ is put equal to zero in (1) after differentiation.

He found that (1) was not amenable to numerical calculations if m is odd. The simplest case is $m = 2$ for which the probability integral reduces to

$$ {}_\nu P_k(Q) = \sum_{r=0}^{k} (-1)^r\, {}^kC_r \left(1 + \frac{2rQ^2}{\nu}\right)^{-\frac{1}{2}\nu}. \tag{3}$$

Hartley (1938) had suggested that if we assume the k variance ratios to be independent, an assumption quite justifiable if ν is very large, ${}_\nu P_k(Q)$ can be obtained approximately from the formula

$$ {}_\nu P_k(Q) \simeq \{{}_\nu P_1(Q)\}^k. \tag{4}$$

* This appears as $1-M$ in equation (10) of Finney's paper through a printing error.

For $m = 2$, this becomes $${}_{\nu}P_k(Q) \simeq \left\{1-\left(1+\frac{2Q^2}{\nu}\right)^{-\frac{1}{2}\nu}\right\}^k. \tag{5}$$

To get some idea of the adequacy of the approximate approach, Finney compared the 5 % points for F_k obtained by the two methods, when $m = 2$. He concluded that the approximation was satisfactory when ν was moderately large in comparison with k, and presumed that the agreement would generally improve with increase of m. The chief uncertainty according to him was, therefore, when $m = 1$.

An alternative method of calculating (1) which is particularly suitable when $m = 1$ is Hartley's expansion of a studentized integral. Thus, using (1) of Part I, the integral itself is

$${}_{\nu}P_k(Q) = 2\Gamma(\tfrac{1}{2}\nu)^{-1}(\tfrac{1}{2}\nu)^{\frac{1}{2}\nu}\int_0^\infty s_0^{\nu-1}e^{-\frac{1}{2}\nu s_0^2}P_k(Qs_0)\,ds_0, \tag{6}$$

where $$P_k(Q) = \left(2\Gamma(\tfrac{1}{2}m)^{-1}(\tfrac{1}{2}m)^{\frac{1}{2}m}\int_0^Q x^{m-1}e^{-\frac{1}{2}mx^2}\,dx\right)^k. \tag{7}$$

It could easily be verified that (6) and (7) lead to the same result as given in (3) when $m = 2$.

In the form given by (6), ${}_{\nu}P_k(Q)$ can be evaluated with sufficient accuracy using expansion (6) of Part I where $P(Q)$ should be replaced by $P_k(Q)$ given in (7). The coefficients $a_0, a_1, a_2, \ldots$, of this expansion will be seen to involve powers and derivatives of the incomplete gamma-function, thus leading to Laguerre functions. When $m = 1$ or 2 these reduce to certain Hermite functions which are fully tabulated in the *British Association Mathematical Tables*, vol. I. In the general case, the Laguerre functions have to be used; these have been tabulated, although less extensively.

In the special case where $m = 1$, we have

$$P_k(Q) = \left(\sqrt{\frac{2}{\pi}}\int_0^Q e^{-\frac{1}{2}x^2}\,dx\right)^k. \tag{8}$$

This can be expressed in terms of the Hermite function

$$Hh_0(x) = \int_x^\infty e^{-\frac{1}{2}x^2}\,dx \tag{9}$$

tabulated in the *British Association Mathematical Tables*, which also give values of

$$Hh_{-n}(x) = \left(-\frac{d}{dx}\right)^n Hh_0(x) \tag{10}$$

for $n = 1$ to 7. These have been used in preparing the tables described in § 3.

When $m = 2$, we have

$$\begin{aligned} P_k(Q) &= (1-e^{-Q^2})^k \\ &= \{1-Hh_{-1}(x)\}^k, \quad \text{where} \quad x = Q\sqrt{2}. \end{aligned} \tag{11}$$

The first six derivatives of the right-hand side of (11) can be obtained from the *British Association Mathematical Tables*. We could therefore proceed with Hartley's method for the case $m = 2$ as well. It seems simpler, however, to use the exact formula (3). Finney has prepared a table of percentage points when $k = 2$ and 3. We shall consider only the case $m = 1$.

3. *Construction of tables*

Writing $P_k(Q)$ as P_k for brevity, we have, for $m = 1$,

$$\left.\begin{aligned}
(\tfrac{1}{2}\pi)^{\frac{1}{2}k} P_k &= \{\sqrt{\tfrac{1}{2}\pi} - Hh_0(Q)\}^k = A^k \text{ (say)},\\
(\tfrac{1}{2}\pi)^{\frac{1}{2}k} P'_k &= kA^{k-1}Hh_{-1},\\
(\tfrac{1}{2}\pi)^{\frac{1}{2}k} P''_k &= k(k-1)A^{k-2}Hh^2_{-1} - kA^{k-1}Hh_{-2},\\
(\tfrac{1}{2}\pi)^{\frac{1}{2}k} P'''_k &= k(k-1)(k-2)A^{k-3}Hh^3_{-1} - 3k(k-1)A^{k-2}Hh_{-1}Hh_{-2} + kA^{k-1}Hh_{-3},\\
(\tfrac{1}{2}\pi)^{\frac{1}{2}k} P^{\mathrm{iv}}_k &= k(k-1)(k-2)(k-3)A^{k-4}Hh^4_{-1} - 6k(k-1)(k-2)A^{k-3}Hh^2_{-1}Hh_{-2}\\
&\quad + k(k-1)A^{k-2}(3Hh^2_{-2} + 4Hh_{-1}Hh_{-3}) - kA^{k-1}Hh_{-4},\\
(\tfrac{1}{2}\pi)^{\frac{1}{2}k} P^{\mathrm{v}}_k &= k(k-1)(k-2)(k-3)(k-4)A^{k-5}Hh^5_{-1} - 10k(k-1)(k-2)(k-3)A^{k-4}Hh^3_{-1}Hh_{-2}\\
&\quad + 5k(k-1)(k-2)A^{k-3}Hh_{-1}(3Hh^2_{-2} + 2Hh_{-1}Hh_{-3})\\
&\quad - 5k(k-1)A^{k-2}(2Hh_{-2}Hh_{-3} + Hh_{-1}Hh_{-4}) + kA^{k-1}Hh_{-5},\\
(\tfrac{1}{2}\pi)^{\frac{1}{2}k} P^{\mathrm{vi}}_k &= k(k-1)(k-2)(k-3)(k-4)(k-5)A^{k-6}Hh^6_{-1}\\
&\quad - 15k(k-1)(k-2)(k-3)(k-4)A^{k-5}Hh^4_{-1}Hh_{-2}\\
&\quad + 5k(k-1)(k-2)(k-3)A^{k-4}(9Hh^2_{-1}Hh^2_{-2} + 4Hh^3_{-1}Hh_{-3})\\
&\quad - 15k(k-1)(k-2)A^{k-3}(Hh^3_{-2} + 4Hh_{-1}Hh_{-2}Hh_{-3} + Hh^2_{-1}Hh_{-4})\\
&\quad + k(k-1)A^{k-2}(10Hh^2_{-3} + 15Hh_{-2}Hh_{-4} + 6Hh_{-1}Hh_{-5}) - kA^{k-1}Hh_{-6}.
\end{aligned}\right\} \tag{12}$$

These six derivatives are sufficient to calculate a_0, a_1, a_2 and a_3 of expansion (6) of Part I. Since a_4 involves Hh_{-8} which is not included in the *British Association Mathematical Tables*, it is not easy to calculate a_4 by the direct method. The calculation of a_3 itself is rather complicated by direct evaluation from the derivatives of P_k and was therefore accomplished by an indirect method of numerical differentiation from a_1 and a_2. The same method could have been used to calculate a_4, but the idea was abandoned as that term does not appear seriously to affect the accuracy, unless k is large and ν is small. A small panel of values of a_4 was, however, calculated by this method for select combinations of values of Q and k to get some idea of its range of magnitude. These are discussed later in this section.

Since values of $Q^rP^{(r)}_k$ were required, an auxiliary function

$$h_i = Hh_{-i}\,Q^iA^{i-1} \tag{13}$$

was introduced in terms of which $Q^rP^{(r)}_k$ became, for $r = 1, 2, 3$ and 4,

$$\left.\begin{aligned}
(\tfrac{1}{2}\pi)^{\frac{1}{2}k} QP' &= kA^{k-1}h_1,\\
(\tfrac{1}{2}\pi)^{\frac{1}{2}k} Q^2P'' &= kA^{k-2}\{(k-1)h_1^2 - h_2\},\\
(\tfrac{1}{2}\pi)^{\frac{1}{2}k} Q^3P''' &= kA^{k-3}\{(k-1)(k-2)h_1^3 - 3(k-1)h_1h_2 + h_3\}\\
(\tfrac{1}{2}\pi)^{\frac{1}{2}k} Q^4P^{\mathrm{iv}} &= kA^{k-4}\{(k-1)(k-2)(k-3)h_1^4 - 6(k-1)(k-2)h_1^2h_2\\
&\quad + (k-1)(3h_2^2 + 4h_1h_3) - h_4\}.
\end{aligned}\right\} \tag{14}$$

Expressions for a_0, a_1 and a_2 are

$$\left.\begin{aligned}
a_0 &= \left(\sqrt{\frac{2}{\pi}}A\right)^k,\\
a_1 &= \tfrac{1}{4}k(\tfrac{1}{2}\pi)^{-\frac{1}{2}k}[\{(k-1)h_1^2 - h_2\}A^{k-2} - h_1A^{k-1}],\\
a_2 &= \tfrac{1}{96}(3Q^4P^{\mathrm{iv}} - 2Q^3P''') - \tfrac{1}{8}a_1.
\end{aligned}\right\} \tag{15}$$

It was decided to limit the values of k to $\leqslant 10$. Using the *British Association Mathematical Tables*, values of A^k, h_1, h_2, h_3 and h_4 were calculated to six decimal accuracy in the range 1·6 (0·2) 6·6 for Q. Values of a_0, a_1 and a_2 were then obtained and all the calculations checked by fifth order differencing on the National Accounting Machine.

The method used to obtain a_3 was to express it in terms of a_1, a_2 and their derivatives a_1', a_2' and a_2''. Thus

$$a_3 = \tfrac{1}{12}Q^2a_2'' - \tfrac{19}{36}Qa_2' + \tfrac{16}{27}a_2 + \tfrac{4}{81}(Qa_1' - 4a_1). \tag{16}$$

The values of a_1', a_2' and a_2'' were obtained by numerical differentiation, which gave the final values of a_3 to two decimal accuracy; this was ample for our purpose. An independent check was provided for a few test values, taking $k = 2$. For obtaining these values, P^{iv}, P^{v} and P^{vi} were calculated and substituted in the formula

$$a_3 = \tfrac{1}{384}(Q^6P^{\text{vi}} + Q^5P^{\text{v}} + 11Q^4P^{\text{iv}}) - \tfrac{3}{2}a_2 - \tfrac{11}{32}a_1 \tag{17}$$

and compared with those obtained with the help of (16). The comparison is set out in Table 1. The agreement is satisfactory for our purpose.

Table 1. *Values of* a_3

Q	Formula (16)	Formula (17)
2·2	1·8190	1·8242
2·4	2·3353	2·3350
2·6	2·3178	2·3176
2·8	1·7453	1·7449
3·0	0·7801	0·7799

A further short-cut was used in calculating a_3. Instead of using (16) for each value of k ranging from 2 to 10, a_3 was first calculated for $k = 2$, 6 and 10. By three-point interpolation between them, values of a_3 were obtained for the remaining values of k.

Although a_4 was not used in the evaluation of ${}_\nu P_k(Q)$, a few of its values were calculated by numerical differentiation of a_1 and a_2, using the formula

$$a_4 = \tfrac{1}{192}Q^4a_2^{\text{iv}} - \tfrac{17}{288}Q^3a_2''' + \tfrac{439}{1728}Q^2a_2'' - \tfrac{9401}{25920}Qa_2' - \tfrac{62}{243}a_2 + \tfrac{44}{3645}(Qa_1' - 4a_1). \tag{18}$$

These values are presented in Table 2.

Table 2. *Values of* a_4

Q \ k	2	6	10
2·8	+ 2·5	+ 0·1	− 60·0
3·6	+ 12·0	+ 50·0	+ 108·9
4·4	− 6·1	− 18·1	− 28·1

The contribution from a_4 to the probability integral may be as large as 0·01 near the upper 5 and 1 % points if $k = 10$, $\nu = 10$. This will affect the percentage point of the largest variance ratio by as much as 1 in the unit's place. If $\nu = 20$, only the first decimal will be affected by about 1 or 2. It is likely in practice that ν will be of the order of 20 or more when $k = 10$, so that the table will not vitiate the level of significance to any serious extent.

The values of ν selected for constructing the tables of the upper 5 and 1 % points of F_k were 10, 12, 15, 20, 30, 60 and ∞ to facilitate harmonic interpolation inside this range. Second-difference inverse interpolation was used to obtain the percentage points from the table of probability integrals.

Table 3 shows the final results giving the upper 5 and 1 % points for $k = 1$ to 10. The values for $k = 1$ were copied from Merrington & Thompson's (1943) tables, and the difference between these and the kth column ($k > 1$) shows how much out we might be in assessing the significance of each variance ratio singly, rather than as one value in a group of k. Each row and column of the percentage points in Table 3 were differenced to the third order as a final check. No suspicious-looking values were found.

Table 3. *Upper per cent points of the largest variance ratio**

5 % points

ν \ k	1	2	3	4	5	6	7	8	9	10
10	4·96	6·79	8·00	8·96	9·78	10·52	11·18	11·79	12·36	12·87
12	4·75	6·44	7·53	8·37	9·06	9·68	10·20	10·68	11·12	11·53
15	4·54	6·12	7·11	7·86	8·47	8·98	9·43	9·82	10·19	10·52
20	4·35	5·81	6·72	7·40	7·94	8·39	8·79	9·13	9·44	9·71
30	4·17	5·52	6·36	6·97	7·46	7·87	8·21	8·51	8·79	9·03
60	4·00	5·25	6·02	6·58	7·02	7·38	7·68	7·96	8·20	8·41
∞	3·84	5·00	5·70	6·21	6·60	6·92	7·20	7·44	7·65	7·84

1 % points

ν \ k	1	2	3	4	5	6	7	8	9	10
10	10·04	13·17	15·08	16·43	17·43	18·25	18·91	19·48	19·97	20·41
12	9·33	11·88	13·52	14·73	15·69	16·47	17·12	17·68	18·16	18·60
15	8·68	10·82	12·18	13·21	14·03	14·72	15·30	15·81	16·26	16·66
20	8·10	9·93	11·08	11·93	12·61	13·19	13·67	14·09	14·49	14·83
30	7·56	9·16	10·14	10·86	11·43	11·90	12·31	12·66	12·97	13·26
60	7·08	8·49	9·34	9·95	10·43	10·82	11·15	11·45	11·72	11·95
∞	6·63	7·88	8·61	9·15	9·54	9·87	10·16	10·41	10·62	10·82

* k is the number of independent variance estimates, each based on 1 degree of freedom. ν denotes the degrees of freedom of the independent 'error' variance.

4. *An illustration*

A good example to illustrate a possible application of Table 3 is provided in Wishart's (1938) paper. He analysed the data (number of sticks) of a uniformity trial on asparagus as if it were an experiment with nitrate (N), phosphate (P) and potash (K) at two levels each, in eight blocks of four plots confounding the three-factor interaction between blocks.† The analysis of variance is set out in Table 4.

† The purpose of this preliminary trial was to collect material for adjustment, by the covariance technique, of inherent fertility differences among the plots when the N, P, K treatments were given in the ensuing season.

Wishart writes: 'The mean square for total "treatments" (6 D.F.) is not significant, but on examination of the separate effects it would appear that the PK interaction is significant at the 5 % level [$F_{0.05} = 4{\cdot}41$, as against the observed $F = 7{\cdot}02$.] With many ordinary experiments we should be happy to claim such a result as indicating a real effect.... But since the treatments were not applied, the effect is entirely accidental.'

He then applies Bartlett's test for homogeneity of variance (which in this case is equivalent to the Neyman-Pearson L_1 test) on the six treatment variances and finds that 'the set of mean squares is compatible with the hypothesis that it is homogeneous, and awkward explanations are avoided'.

D. J. Bishop & U. S. Nair (1939) found that Bartlett's test underestimates the significance if the degrees of freedom of some of the k variances are as small as 1 or 2. The 5 % point of L_1 which they give when $k = 6$ and $\nu = 1$ is 0·094 against Bartlett's value of 0·077. The value of L_1 for the present example is 0·152. It is not significant,* even by the exact L_1-test of Bishop & Nair.

Table 4. *Analysis of variance*

Variation	Degrees of freedom	Mean square (variance)	Variance ratio (F)
Blocks	7	6543·2	—
N	1	488·3	0·35
P	1	69·0	0·05
K	1	34·0	0·02
NP	1	830·3	0·60
NK	1	57·8	0·04
PK	1	9765·0	7·02
(Total treatments)	(6)	(1874·1)	—
Error	18	1391·2	—

Since only *one* of the six variances is standing out prominently, Cochran's (1941) test based on g, the ratio of the largest variance divided by the sum of the set of variances may appeal as more appropriate than the L_1-test. The value of g for the six treatment variances of Table 4 is 0·87. The 5 % value of g given in Cochran's table is 0·78. According to the g-test, therefore, the variance for interaction PK is significantly large, making us reject a hypothesis, which we know to be true.

To decide which of the two tests will be more appropriate in a situation where the heterogeneity among the k variances is caused by the presence of a single variance much larger than the rest, would need a study of the power functions of the two tests for this particular class of alternative hypotheses, using the Neyman-Pearson theory.

In the present example, however, neither the L_1 nor the g-test takes into account the estimated 'error' variance with the help of which it has been possible to say that the total treatment effects (6 D.F.) were *not* significant. The question now to be answered is whether the variance due to interaction PK, being the largest of the six individual treatment effects, could be considered as significantly greater than the error variance. The variance ratio for PK is 7·02. The 5 % point in Table 3 for $k = 6$, $\nu = 18$ is 8·59. The observed largest variance ratio is, therefore, not significantly large at the 5 % level.

* For L_1, significance is indicated by *low*, not high, values.

5. *Probability integral of the smallest variance ratio*

Let ${}_\nu p_k(q)$ denote the probability that s_1/s_0 is $\geqslant q$. The probability that the smallest variance ratio F_1 is $\geqslant q^2$ will also be given by ${}_\nu p_k(q)$.

Finney (1941) showed that
$$ {}_\nu p_k(q) = (1-M)^k (1+\lambda)^{-\frac{1}{2}\nu}, \tag{19}$$
where M is as defined in (2) except for a replacement of Q by q; and λ is to be put equal to zero after differentiation.

Using Hartley's general studentized integral
$$ {}_\nu p_k(q) = 2\Gamma(\tfrac{1}{2}\nu)^{-1}(\tfrac{1}{2}\nu)^{\frac{1}{2}\nu}\int_0^\infty s_0^{\nu-1}e^{-\frac{1}{2}\nu s_0^2}p_k(qs_0)\,ds_0, \tag{20}$$
where
$$ p_k(q) = \left(2\Gamma(\tfrac{1}{2}m)^{-1}(\tfrac{1}{2}m)^{\frac{1}{2}m}\int_q^\infty x^{m-1}e^{-\frac{1}{2}mx^2}\,dx\right)^k. \tag{21}$$

Finney remarks: 'The convergence of the series by which significance levels for the smallest variance ratio are obtained is much more rapid than for the largest ratio. Also, as Hartley has demonstrated, the approximation given by an assumption of independence is very close even for small numbers of degrees of freedom.'

These remarks could easily be substantiated when $m = 2$ for which the exact value of ${}_\nu p_k(q)$ obtained from either (19) or (20) is
$$ \left(1+\frac{2kq^2}{\nu}\right)^{-\frac{1}{2}\nu}. \tag{22}$$

On the assumption of independence as in (4), the approximate value for ${}_\nu p_k(q)$ will be
$$ \left(1+\frac{2q^2}{\nu}\right)^{-\frac{1}{2}k\nu}. \tag{23}$$

Evidently (22) and (23) are not identical,* but when q is small and ν is not too small, they will be very close to each other.

As we are concerned with the lower percentage points of the smallest variance ratio, q will be small and hence approximation (4) should yield sufficiently accurate results for any m.

We may, however, examine this more closely for the case $m = 1$. The exact value of ${}_\nu p_k(q)$ can be obtained using the expansion (6) of Part I. Owing to the smallness of q, it is scarcely necessary to add terms beyond a_1. We may, however, include a_2 which is approximately equal to $-a_1/8$. The a_0 and a_1 terms are calculated from the formulae
$$ a_0 = p_k(q) = \left(\sqrt{\frac{2}{\pi}}\int_q^\infty e^{-\frac{1}{2}x^2}\,dx\right)^k = (1-\alpha)^k, \tag{24}$$
$$ a_1 = k(qz)(1-\alpha)^{k-2}\{(k-1)qz+\tfrac{1}{2}(q^2+1)(1-\alpha)\}, \tag{25}$$
where α and z are quantities given in Table 2 of the *Tables for Statisticians and Biometricians*, Part I.

In Table 5, values of a_0 and a_1 are calculated for q in the range 0 to 0·1 at intervals of 0·01 and for $k = 1$ to 10. The probability that the smallest of k variance ratios, F_1 is $\leqslant q^2$ will be given by
$$ 1-{}_\nu p_k(q) = \left\{1-a_0-\frac{a_1}{\nu}\left(1-\frac{1}{8\nu}\right)\right\}, \tag{26}$$

* Finney claims in equation (21) of his paper that they are identical. obviously by oversight.

Table 5. *Table for calculating the probability integral of the smallest variance ratio**

k / q	2		3		4	
	a_0	a_1	a_0	a_1	a_0	a_1
0·00	1·000 000	0·000 00	1·000 000	0·000 00	1·000 000	0·000 00
0·01	0·984 106	0·003 99	0·976 254	0·005 98	0·968 465	0·007 98
0·02	0·968 341	0·007 98	0·952 890	0·011 97	0·937 685	0·015 95
0·03	0·952 707	0·011 97	0·929 906	0·017 95	0·907 650	0·023 90
0·04	0·937 204	0·015 97	0·907 301	0·023 93	0·878 352	0·031 84
0·05	0·921 835	0·019 97	0·885 074	0·029 90	0·849 780	0·039 74
0·06	0·906 600	0·023 97	0·863 225	0·035 87	0·821 924	0·047 61
0·07	0·891 502	0·027 98	0·841 750	0·041 83	0·794 775	0·055 43
0·08	0·876 540	0·032 00	0·820 649	0·047 78	0·768 323	0·063 20
0·09	0·861 717	0·036 02	0·799 921	0·053 72	0·742 556	0·070 90
0·10	0·847 034	0·040 05	0·779 563	0·059 64	0·717 466	0·078 53

k / q	5		6		7	
	a_0	a_1	a_0	a_1	a_0	a_1
0·00	1·000 000	0·000 00	1·000 000	0·000 00	1·000 000	0·000 00
0·01	0·960 738	0·009 97	0·953 072	0·011 96	0·945 468	0·013 95
0·02	0·922 723	0·019 92	0·907 999	0·023 88	0·893 511	0·027 83
0·03	0·885 927	0·029 83	0·864 725	0·035 72	0·844 029	0·041 56
0·04	0·850 327	0·039 68	0·823 196	0·047 44	0·796 930	0·055 09
0·05	0·815 893	0·049 46	0·783 357	0·059 00	0·752 118	0·068 36
0·06	0·782 600	0·059 13	0·745 157	0·070 38	0·709 505	0·081 30
0·07	0·750 421	0·068 69	0·708 543	0·081 53	0·669 002	0·093 88
0·08	0·719 332	0·078 11	0·673 465	0·092 42	0·630 523	0·106 05
0·09	0·689 306	0·087 38	0·639 874	0·103 03	0·593 986	0·117 76
0·10	0·660 316	0·096 48	0·607 718	0·113 34	0·559 310	0·128 98

k / q	8		9		10	
	a_0	a_1	a_0	a_1	a_0	a_1
0·00	1·000 000	0·000 00	1·000 000	0·000 00	1·000 000	0·000 00
0·01	0·937 924	0·015 94	0·930 440	0·017 92	0·923 017	0·019 90
0·02	0·879 253	0·031 76	0·865 223	0·035 67	0·851 417	0·039 56
0·03	0·823 829	0·047 35	0·804 112	0·053 09	0·784 867	0·058 75
0·04	0·771 503	0·062 63	0·746 886	0·070 03	0·723 056	0·077 29
0·05	0·722 126	0·077 50	0·693 329	0·086 39	0·665 681	0·095 03
0·06	0·675 559	0·091 88	0·643 237	0·102 06	0·612 462	0·111 83
0·07	0·631 667	0·105 70	0·596 416	0·116 95	0·563 132	0·127 60
0·08	0·590 320	0·118 92	0·552 679	0·131 00	0·517 439	0·142 25
0·09	0·551 390	0·131 47	0·511 848	0·144 14	0·475 142	0·155 72
0·10	0·514 758	0·143 32	0·473 755	0·156 32	0·436 017	0·167 97

$$1 - {}_{\nu}p_k(q) = 1 - a_0 - \frac{a_1}{\nu}\left(1 - \frac{1}{8\nu}\right),$$ see equation (26).

which can be calculated with the help of Table 5 for any given q, k and ν. If this probability is less than 0·05 or 0·01, the observed smallest variance ratio q^2 can be said to be significantly small at the 5 or 1 % level respectively.

The values of ${}_\nu p_k(q)$ given by Table 5 have been compared in Table 6 with those obtained by the approximate formula (4), for $\nu = 10$, $q = 0·10$ and 0·01.

Table 6. *Comparison of exact and approximate values of the probability integrals* of the smallest variance ratio*

k	$q = 0·10$		$q = 0·01$	
	${}_{10}p_k(q)$	$\{{}_{10}p_1(q)\}^k$	${}_{10}p_k(q)$	$\{{}_{10}p_1(q)\}^k$
1	0·922 324	0·922 321	0·992 218	0·992 218
2	0·850 989	0·850 676	0·984 500	0·984 497
3	0·785 452	0·784 596	0·976 845	0·976 836
4	0·725 221	0·723 649	0·969 253	0·969 234
5	0·669 843	0·667 437	0·961 722	0·961 691
6	0·618 911	0·615 591	0·954 253	0·954 207
7	0·572 047	0·567 773	0·946 845	0·946 781
8	0·528 911	0·523 669	0·939 498	0·939 413
9	0·489 191	0·482 990	0·932 210	0·932 102
10	0·452 604	0·445 472	0·924 982	0·924 848

There is satisfactory agreement between the two methods. Indeed, as is to be expected, the agreement is much better for $q = 0·01$ than for $q = 0·10$. The difference in the top row values in columns (2) and (3) is due to calculating the latter up to the term in ν^{-4} using the incomplete beta-function expansion (10) of Part I. It will be seen that for this expansion, terms up to ν^{-2} give a five decimal accuracy when $q = 0·10$ and this improves to six decimals when $q = 0·01$.

We may, before concluding, add an illustration for the use of the test of the smallest variance ratio. This is also taken from Wishart's (1938) paper, where by the ordinary test of significance he gets variance for an interaction NP as significantly less than the error variance. He calls this a '*negative* interaction'. Applying Bartlett's test, Wishart found that this 'negative' interaction was not significant. The values concerned are $s_1^2 = 0·01$, $s_0^2 = 5·57$, $k = 6$, $m = 1$, $\nu = 18$. To apply the test for significance of the smallest variance ratio, we calculate $F_1 = \dfrac{0·01}{5·57} = 0·0018$ giving $q = 0·04$. Using (26) and referring to Table 5, we get

$${}_{18}p_6(0·04) = a_0 + a_1(1 - 8/\nu)/\nu = 0·8258.$$

The probability that F_1 is less than or equal to the observed value is about 0·17, which is much greater than 0·05. The observed F_1 is therefore not significantly small.

In conclusion, I should like to acknowledge warmly the help and guidance I have received from Prof. E. S. Pearson and Dr H. O. Hartley in the course of my investigations.

* The probability integral is calculated here from q to ∞, instead of the usual limits 0 to q.

REFERENCES

BISHOP, D. J. & NAIR, U. S. (1939). A note on certain methods of testing for the homogeneity of a set of estimated variances. *J. Roy. Statist. Soc. Suppl.* **6**, 89.

COCHRAN, W. G. (1941). The distribution of the largest of a set of estimated variances as a fraction of their total. *Ann. Eugen., Lond.*, **11**, 47.

FINNEY, D. J. (1941). The joint distribution of variance ratios based on a common error mean square. *Ann. Eugen., Lond.*, **11**, 136.

FISHER, R. A. (1924). On a distribution yielding the error functions of several well-known statistics. *Proc. Intern. Math. Congress*, Toronto, pp. 805–13.

FISHER, R. A. (1926). Expansion of 'Student's' integral in powers of n^{-1}. *Metron*, **5**, 109.

HARTLEY, H. O. (1938). Studentization and large-sample theory. *J. Roy. Statist. Soc. Suppl.* **5**, 80.

HARTLEY, H. O. (1944). Studentization, or the elimination of the standard deviation of the parent population from the random sample-distribution of statistics. *Biometrika*, **33**, 173.

MAHALANOBIS, P. C. (1944). On large-scale sample surveys. *Phil. Trans. Roy. Soc.* B, **231**, 329.

MERRINGTON, M. & THOMPSON, C. M. (1943). Tables of percentage points of the inverted Beta (F) distribution. *Biometrika*, **33**, 73.

NAIR, K. R. (1948). The distribution of the extreme deviate from the sample mean and its studentized form. *Biometrika*, **35**, 118.

NEYMAN, J. & PEARSON, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, **20** A, 175, 263.

PEARSON, E. S. & HARTLEY, H. O. (1943). Tables of the probability integral of the 'studentized' range. *Biometrika*, **33**, 89.

'STUDENT' (1908). On the probable error of a mean. *Biometrika*, **6**, 1.

SUKHATME, P. V. (1937). Tests of significance for samples of the χ^2-population with two degrees of freedom. *Ann. Eugen., Lond.*, **8**, 52.

THOMPSON, C. M. (1941). Tables of percentage points of the incomplete beta-function. *Biometrika*, **32**, 151.

WISHART, J. (1927). On the approximate quadrature of certain skew curves with an account of the researches of Thomas Bayes. *Biometrika*, **19**, 1.

WISHART, J. (1938). Field experiments of factorial design. *J. Agric. Sci.* **28**, 299.

THE ESTIMATION OF NON-LINEAR PARAMETERS BY 'INTERNAL LEAST SQUARES'

By H. O. HARTLEY

1. Introduction

(1·1) *The difficulties arising with curvilinear regressions*

In statistical regression of two variables y and x, the linear relation $y = a+bx$ has received most attention. Apart from occasional 'fitting of polynomials', non-linear relations have not been used frequently. As reasons for the reluctance on the part of statistical workers to use non-linear regressions we may mention here:

(*a*) The complication in the computational procedure when estimating non-linear parameters by efficient methods such as least squares.

(*b*) The lack of exactness in tests of the goodness of fit and the difficulty of establishing the random sampling distribution of fitted statistics.

(*c*) The fact that any transformations to linearity such as $f(y) = bg(x)+a$ (which may be suggested by theory), usually involve certain unknown parameters in f and g which must be estimated from the sample and, therefore, bring in difficulties of type (*b*).

(*d*) The difficulty of deciding which of the many possible non-linear regressions is suggested by theory.

As examples illustrating these difficulties we may mention here:

(*a*) The iterative process in dosage mortality technique which necessitates the preparation of special auxiliary tables for each fitted law.

(*b*) Methods of fitting the Gompertz curve by factorial moments (Sasuly, 1934) resulting in procedures of unknown efficiency.

(*c*) The exponential law of diminishing returns:

$$y = \hat{y}(1-e^{kx}) \tag{1}$$

can be transformed to a linear relation between $\log_e(1-y/\hat{y})$ and x:

$$\log_e(1-y/\hat{y}) = kx. \tag{1'}$$

However, a knowledge of the 'limiting response' $\hat{y}$ is usually required in order to make this transformation to linearity. If $\hat{y}$ is assessed by 'inspection' the accuracy of the subsequent fit of k cannot be ascertained, let alone the accuracy of the assessment of $\hat{y}$.

(*d*) In numerous applications polynomial regressions have been fitted without a biological or social theory to guide the fit and, as a result, difficulties have been encountered in interpreting the real meaning of the polynomial terms.

(1·2) *The principle of 'internal regression'*

The principle here adopted is that of bringing most of the important regression laws under one generating principle: Most of the laws in physics and technology are generated by simple, mainly linear, relationships, between the function and its first and second derivative (e.g. velocity and acceleration) and the same applies to most of the important laws in the biological sciences. We may name two examples:

(A) The exponential law mentioned above is generated by the first order differential equation

$$\frac{dy}{dx} = -k\hat{y} + ky, \tag{2}$$

which is a linear relationship. If applied to fertilizer response in agriculture ($k < 0$), this has the simple meaning that the additional yield, dy, caused by an addition of fertilizer dx, is, in the first place, proportional to dx (i.e. $= \hat{y}(-k)\,dx$); but this constant rate of increase ($\hat{y}(-k)$) is retarded by an amount $(-(-k)y)$ which in turn is proportional to the yield y already attained.

(B) The 'trade cycle' for, say, a production index y is sometimes regarded as being caused by the interplay of production y and demand z which are assumed to satisfy the relations

$$\frac{dy}{dt} = \alpha z, \quad \frac{dz}{dt} = -\beta y,$$

resulting in

$$\frac{d^2y}{dt^2} = -\alpha\beta y \tag{3}$$

of which the general solution is the trade cycle law:

$$y = \sin\left[\sqrt{(\alpha\beta)}\,t + \gamma\right]. \tag{4}$$

The idea, now, is to fit directly to the data the generating *linear* law (such as (2) or (3)) resulting in linear equations for the parameter estimates, rather than fitting the non-linear regressions (such as (1) or (4)). However, some modications are required: whilst for a mathematical function $y(x)$ the regression (1) and the differential equation (2) are identical conditions, it would be difficult to fit to an empirical series of observed y_i a condition involving its differential coefficient. We therefore seek a finite difference equivalent to the differential equation (2). As is well known (see, for example, Bartlett, 1946; Cunningham & Hynd, 1946; Kendall, 1944) the first order linear difference equation of the form

$$y_i - y_{i-1} = by_i + a \tag{5}$$

is capable of generating exactly the exponential law of type (1). Integrating (or summing) (5) we obtain

$$y_j = bY_j + ax_j + c, \tag{6}$$

where $Y_j = \sum_{i=0}^{j} y_i$. Now this equation may be regarded as a linear regression equation for the dependent variable y_j with its own progressive sum Y_j as independent variable, and with x_j playing the part of a second independent variable. Formally, therefore, the non-linear parameters $\hat{y}$ and k of (1) can be estimated from the simple linear regression coefficients a, b, c given in (6).†

In a similar manner the second order differential equation (3) can be replaced by a linear relation between y_j and its second sum, ${}_2Y_j = \sum_{l=0}^{j}\sum_{i=0}^{l} y_i$, and such a relation can be shown to be equivalent to the harmonic regression law (trade cycle) (4). The period $\sqrt{(\alpha\beta)}$ of this trade cycle can therefore be determined as a partial linear regression coefficient between y_j and ${}_2Y_j$.

Turning now to the general case, we shall call any regression equation in which the dependent variate y is related to its own repeated sums

$${}_1Y_j = \sum_{i=0}^{j} y_i, \quad {}_2Y_j = \sum_{l=0}^{j}\sum_{i=0}^{l} y_i \tag{7}$$

† In § (2·1) we shall give the exact relation between the parameters y_0, k and a, b, c.

etc., as independent variables, an 'Internal regression'. This concept is, of course, closely linked with that of auto-regression in time series. Its usefulness lies in its close relation to linear differential equations, which are known to generate most of the important regression laws and also give a better understanding of the physical or biological mechanism producing the respective curvilinear regression.

The principle of estimation and goodness of fit that we shall use in conjunction with internal regression is that of least squares applied to the 'integrated' equation (6) and not to the corresponding difference equation (5) (see, for example, Bartlett, 1946; Mann & Wald, 1943). This method is, at first, accepted as a working rule and the formulae for the estimators derived and demonstrated in terms of examples. Only the first order internal regression is treated here (§ 2). The efficiency of the estimator is then considered under various assumptions of residual independence. In particular, the classical assumption of random residuals attached to the observed y_i is fully investigated by comparing the efficiency of our estimators with the 100 % efficient transcendental maximum likelihood roots. To this end, large sample formulae for the variances of the estimators will be developed (§ 3), leaving the derivation of exact sampling distributions for goodness of fit tests to a later paper.

2. The procedure of internal regression

(2·1) *The first order internal regression; fitting of the exponential law*

Consider a sample of observed values y_i corresponding to discrete integer positions, $x_i = i$, of the independent variable x (e.g. a time series). If the y_i satisfy exactly the difference equation

$$(y_{i+1}-y_i) = -\tfrac{1}{2}b(y_{i+1}+y_i)+a \tag{8}$$

then, by standard finite difference technique, they also satisfy exactly the exponential law

$$y_i = \hat{y}(1-fe^{kx_i}), \tag{9}$$

where the limiting response $\hat{y}$ and the 'exponential curvature' k are given by

$$\hat{y} = a/b, \quad -k = 2\tanh^{-1}\tfrac{1}{2}b \tag{10}$$

whilst f is a constant of integration. In order to find the least square estimates for a and b it is convenient to distinguish two cases:

(2·11) $n = \text{odd} = 2m+1 \quad (i = -m, -m+1, \ldots, -1, 0, 1, \ldots, m)$,

(2·12) $n = \text{even} = 2m \quad (i = -m, \ldots, -1, 1, \ldots, m)$.

(2·11) We introduce

$$\bar{y} = (2m+1)^{-1}\sum_{i=-m}^{+m} y_i, \quad \eta_i = y_i-\bar{y}, \quad \xi_i = i. \tag{11}$$

Hence

$$\Sigma\eta_i = \Sigma\xi_i = 0. \tag{12}$$

We rewrite (8) as

$$(\eta_{i+1}-\eta_i) = -\tfrac{1}{2}b(\eta_{i+1}+\eta_i)+(a-b\bar{y}) \tag{13}$$

and sum the difference equation form $i = 0$ to $i = j$, introducing at the same time as a new variable S_j the first sum of the η_i as follows:

$$\begin{aligned} S_j &= \tfrac{1}{2}\eta_0+\sum_{i=1}^{j-1}\eta_i+\tfrac{1}{2}\eta_j \quad \text{for} \quad j\geqslant 1,\\ S_0 &= 0,\\ S_j &= -\tfrac{1}{2}\eta_0-\sum_{i=-1}^{j+1}\eta_i-\tfrac{1}{2}\eta_j \quad \text{for} \quad j\leqslant -1. \end{aligned} \tag{14}$$

The progressive sums of (13) are then found to be equivalent to the equations:

$$\eta_j - \eta_0 = -bS_j + (a - b\bar{y})\,\xi_j + c, \tag{15}$$

or

$$\eta_j = -bS_j + a'\xi_j + c', \tag{16}$$

where

$$a' = a - b\bar{y},$$

or

$$a'/b = y_0 - \bar{y} \tag{17}$$

and c, c' are constants of summation.

The least square solutions for a', $-b$ and c' are then given by:

$$\left.\begin{aligned} -b\Sigma S_i^2 + a'\Sigma S_i\xi_i + c'\Sigma S_i &= \Sigma S_i\eta_i, \\ -b\Sigma S_i\xi_i + a'\Sigma\xi_i^2 \qquad\qquad &= \Sigma\eta_i\xi_i, \\ -b\Sigma S_i \qquad\qquad + c'n &= 0. \end{aligned}\right\} \tag{18}$$

It is now easy to see from the definition of the S_j that they are exactly orthogonal to the η_j, i.e. that

$$\sum_{i=-m}^{+m} S_i\eta_i = 0. \tag{19}$$

Using (19), subtracting the last equation in (18) from the first two and solving for $-b$ and a' we find

$$-b = \Delta^{-1}(\Sigma\eta_i\xi_i\Sigma S_i\xi_i), \quad a' = \Delta^{-1}(\Sigma(S_i - \bar{S})^2\,\Sigma\eta_i\xi_i), \tag{20}$$

where

$$\Delta = \Sigma(S_i - \bar{S})^2\,\Sigma\xi_i^2 - (\Sigma S_i\xi_i)^2, \quad \bar{S} = n^{-1}\Sigma S_i$$

and the summation is extended from $i = -m$ to $i = m$ throughout. For practical use the formulae (20) must be further simplified; in particular, we wish to avoid the calculation of the deviates η_i. We therefore introduce sum values formed directly from the y_i

$$\left.\begin{aligned} Y_j &= \tfrac{1}{2}y_0 + \sum_{i=1}^{j-1} y_i + \tfrac{1}{2}y_j \quad \text{for} \quad j \geq 1, \\ Y_0 &= 0, \\ Y_j &= -\tfrac{1}{2}y_0 - \sum_{i=-1}^{j+1} y_i - \tfrac{1}{2}y_j \quad \text{for} \quad j \leq -1, \\ Y &= \Sigma y_i, \end{aligned}\right. \tag{21}$$

and note the following relationships:

$$\Sigma\xi_i S_i = -\tfrac{1}{2}\Sigma\eta_i\xi_i^2, \tag{22}$$

$$\Sigma(S_i - \bar{S})^2 = \Sigma Y_i^2 + Yn^{-1}\Sigma\eta_i\xi_i^2 - Y^2n^{-2}\Sigma\xi_i^2 - n^{-1}(\Sigma Y_i)^2. \tag{23}$$

$$\Sigma S_i = \Sigma Y_i. \tag{24}$$

With the help of these relations we reach the following working formulae for the estimators a', $-b$ and c' in the order of computation. From data calculate Y_i, ΣY_i and Y (see (21)),

$$\Sigma y_i\xi_i, \quad \Sigma y_i\xi_i^2, \quad \Sigma Y_i^2, \quad \Sigma\xi_i^2 = \tfrac{1}{3}m(m+1)(2m+1).$$

Compute

$$\left.\begin{aligned} &\Sigma\eta_i\xi_i^2 = \Sigma y_i\xi_i^2 - Yn^{-1}\Sigma\xi_i^2, \\ &\Sigma(S_i - \bar{S})^2 \text{ from (23)}, \\ &\Delta = \Sigma\xi_i^2\Sigma(S - \bar{S})^2 - \tfrac{1}{4}(\Sigma\eta_i\xi_i^2)^2, \\ &a' = \Delta^{-1}\Sigma y_i\xi_i\Sigma(S - \bar{S})^2, \\ &-b = \tfrac{1}{2}\Delta^{-1}\Sigma y_i\xi_i\Sigma\eta_i\xi_i^2, \\ &c' = n^{-1}b\Sigma Y_i. \end{aligned}\right. \tag{25}$$

Finally, we note the relation between the estimators a', $-b$ and c' and the parameters of the original exponential law, i.e. the limiting response $\hat{y}$, the exponential curvature k and f. We have
$$\hat{y} = a'/b + n^{-1}Y, \quad f\hat{y} = a'/b - c', \quad k = -2\tanh^{-1}\tfrac{1}{2}b. \tag{26}$$
As an illustration and check of these formulae we apply them in example 1 below to the theoretical sequence $y_i = 1 - e^{-x_i}$, $x_i = 0(1)4$. The table below is self explanatory, the work following the computational order of equations (25).

Example 1. *Theoretical series $y_i = 1 - e^{-x_i}$; estimation of parameters by internal least square*

x_i	y_i	$2Y_i$ (from (21))	ξ_i	ξ_i^2
0	0·000	−2·129	−2	4
1	0·632	−1·497	−1	1
2	0·865	0	0	0
3	0·950	1·815	1	1
4	0·982	3·747	2	4

$Y = 3{\cdot}429$ $\qquad$ $1{\cdot}936 = 2\Sigma Y_i$

$$\begin{array}{llll}
\Sigma y_i\xi_i = 2{\cdot}282, & \Sigma\xi_i^2 = 10, & \Sigma Y_i^2 = 6{\cdot}0270, \\
\Sigma y_i\xi_i^2 = 5{\cdot}510, & & \\
\Sigma\eta_i\xi_i^2 = -1{\cdot}350, & \Sigma(S_i-\bar{S})^2 = 0{\cdot}2119, & \Delta = 1{\cdot}665, \\
a' = 0{\cdot}290, & b = 0{\cdot}925, & c' = 0{\cdot}179, \\
\hat{y} = 1{\cdot}000, & k = -1{\cdot}000, & f\hat{y} = 0{\cdot}135 = e^{-2}.
\end{array}$$

The true values of the parameters are, of course, exactly reproduced.

(2·12) *The case n = even = $2m$*

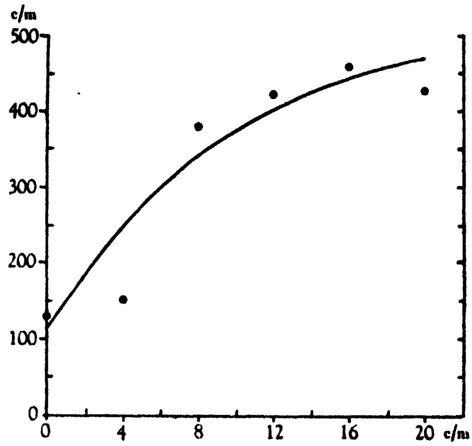
The only modifications of the previous section consist in defining
$$\xi_i = \tfrac{1}{2}(2i-1) \quad \text{for} \quad i = 1, \ldots, m; \qquad \xi_i = \tfrac{1}{2}(2i+1) \quad \text{for} \quad i = -1, \ldots, -m \tag{27}$$
and
$$\begin{aligned} Y_j &= \sum_{i=1}^{j-1} y_i + \tfrac{1}{2}y_j \quad \text{for} \quad j \geqslant 1, \\ Y_j &= -\sum_{i=-1}^{j+1} y_i - \tfrac{1}{2}y_j \quad \text{for} \quad j \leqslant -1, \end{aligned} \tag{28}$$
and similarly for the S_j. All other definitions, formulae and results are the same as for odd n except that the summations are from $i = -m$ to -1 and from $+1$ to $+m$, omitting 0.

Below we illustrate the procedure by fitting the exponential law of diminishing returns to yield data obtained in manurial trials of the National Agricultural Research Bureau, China.† Both trials give the response of a number of wheat varieties to the application of sulphate of ammonia at the rates of 4, 8, 12, 16 and 20 c./m.‡ The responses are grain yields and we confine ourselves here to the mean varietal responses. In the first experiment (example 2) it will be seen that there is a marked retardation of the response to the higher

† The data were kindly put at my disposal by Dr H. L. Richardson of the Imperial Chemical Industries Ltd. Many authors have applied the exponential law to the analysis of fertilizer trials (see, for example, Crowther & Yates (1941)).

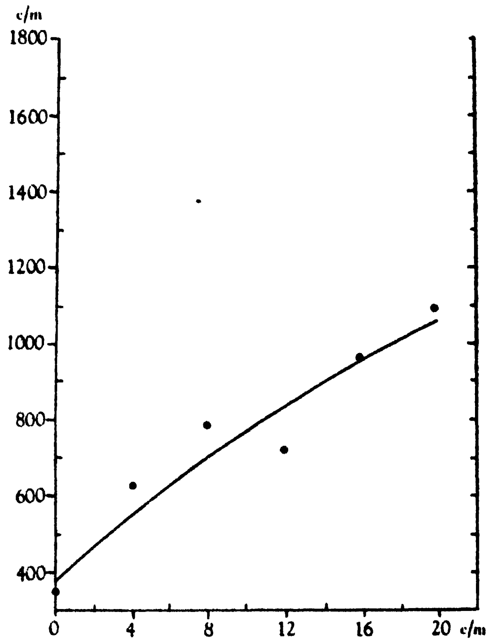
‡ c./m. = cattys per mou, the Chinese measure of yield and rate of fertilizer application.

fertilizer rates. In the second experiment (example 3), there is hardly any retardation and the estimate of the limiting response is, consequently, very high. The fit is illustrated in Figs. 1 and 2.



limiting response, $\hat{y} = 530$ c/m.

Fig. 1. Wheat yield; response to rate of fertilizer (example 2).



limiting response, $\hat{y} = 1786$ c/m.

Fig. 2. Wheat yield; response to rate of fertilizer (example 3).

Normally one would, of course, require more than six observations for fitting a three parameter regression. However, the estimation of exponential curvature and limiting response in short series is akin to the calculation of the 'quadratic effect' customary in fertilizer trials with only three levels.

Example 2. N-*fertilizer trial with wheat, Hopen, Tinghsien*, 1936. *Fit of exponential law of diminishing returns to mean varietal responses*

Fertilizer rate in c./m.	Response $y_i = 10 \times$ yield -1500	$2Y_i$	$2\xi_i$	$4\xi_i^2$
0	127	−1187	−5	25
4	151	−909	−3	9
8	379	−379	−1	1
12	421	421	1	1
16	460	1302	3	9
20	426	2188	5	25

$Y = 1964$ 1436   Check:† $1436 = 6(1307-657)-2(1232)$

$Y^- = 657$*

$Y^+ = 1307$

$$\Sigma y_i\xi_i = 1232, \quad \Sigma\xi^2 = 17{\cdot}5, \quad \Sigma Y_i^2 = 225\,9670,$$
$$\Sigma y_i\xi_i^2 = 5031, \quad \Sigma Y_i = 718,$$
$$\Sigma\eta_i\xi_i^2 = -697, \quad \Sigma(S_i-\bar{S})^2 = 70524, \quad \Delta = 111\,2718,$$
$$a' = 78{\cdot}1, \quad b = +0{\cdot}386, \quad c' = 462,$$
$$\hat{y} = 530, \quad k = -0{\cdot}391, \quad \hat{y}f = 156.$$

Fitted law:
$$y_i = 529{\cdot}6 - 156{\cdot}1\exp(-0{\cdot}391\xi_i).$$

Example 3. N-*fertilizer trial with wheat, Shantung*, 1936. *Fit of exponential law of diminishing returns to mean varietal responses*

Fertilizer rate in c./m.	Response $y_i = 10$ yield -3000	$2Y_i$	$2\xi_i$	$4\xi_i^2$
0	353	−3179	−5	25
4	627	−2199	−3	9
8	786	−786	−1	1
12	717	717	1	1
16	959	2393	3	9
20	1093	4445	5	25

$Y = 4535$ 1391   Check: $1391 = 6(2769-1766)-2(2313{\cdot}5)$

$Y^- = 1766$

$Y^+ = 2769$

$$\Sigma y_i\xi_i = 2313{\cdot}5, \quad \Sigma\xi_i^2 = 17{\cdot}5, \quad \Sigma Y_i^2 = 103\,89\,500,$$
$$\Sigma y_i\xi_i^2 = 12982, \quad \Sigma Y_i = 695{\cdot}5,$$
$$\Sigma\eta_i\xi_i^2 = -245, \quad \Sigma(S_i-\bar{S})^2 = 126230, \quad \Delta = 219\,4019,$$
$$a' = 133, \quad b = 0{\cdot}1292, \quad c' = 15{\cdot}0,$$
$$\hat{y} = 1786, \quad k = -0{\cdot}1294, \quad \hat{y}f = 1015.$$

Fitted law:
$$y_i = 1786 - 1015\exp(-0{\cdot}1294\xi_i).$$

† It is convenient, when adding the y_i, to record $Y^- = \Sigma y_i$ for $i<0$ and $Y^+ = \Sigma y_i$ for $i>0$, separately (forming $Y = Y^- + Y^+ = \Sigma y_i$). This provides a check on the forming and copying of the Y_i from the equation $\Sigma 2Y_i = n(Y^+ - Y^-) - 2\Sigma y_i\xi_i$, e.g. in the present case $6(1307-657)-2(1232) = 1436$.

We now turn to two well-known regression laws which, by suitable transformation, can be reduced to our exponential law.

(2·2) *The internal least square fit of the logistic curve*

The general form of the logistic curve is†

$$z = A/(1+Be^{-kx}). \tag{29}$$

In the case where A is known (as for instance with a biological response known to vary between 0 and 100 %) the transformation

$$\log(A/z-1) = \log B - kx, \tag{30}$$

reduces the problem to a plain linear regression fit for $\log B$ and k. Nevertheless, this case has recently been dealt with (Finney, 1947) by the more elaborate, but under certain conditions more efficient, method of maximum likelihood.

We are here dealing with the general case where A has to be estimated from the data. In this case the transformation $z = 1/y$ yields

$$y = 1/A + (B/A)e^{-kx}; \tag{31}$$

which is of the form (1) or (9). Considerations of appropriateness and efficiency are, again, postponed for discussion in § 3 below, but we should mention here that Rhodes (1940) also uses reciprocals when fitting the logistic by a different method. We are giving in example 4, below, the data used by him, viz. Population figures for the U.S. Census Data, 1800–1910. It is with such data on population growth that the estimation of the logistic parameters is of importance.

Example 4. *U.S. Census population* 1800–1910. *Fit of logistic curve and estimation of its parameters by internal least squares*

Year x	Population (millions) z	$y = 10{,}000/z$	$2Y_t$	$2\xi_t$	$4\xi_t^2$
1800	5·308	1884	−10310	−11	121
10	7·240	1381	−7045	−9	81
20	9·638	1038	−4626	−7	49
30	12·866	777	−2811	−5	25
40	17·069	586	−1448	−3	9
1850	23·192	431	−431	−1	1
60	31·443	318	318	1	1
70	38·558	259	895	3	9
80	50·156	199	1353	5	25
90	62·948	159	1711	7	49
1900	75·995	132	2002	9	81
10	91·972	109	2243	11	121

$$Y = 7273$$
$$Y^- = 6097$$
$$Y^+ = 1176$$

† In some applications it has been found necessary to generalize the logistic to

$$y - C = A/(1+B\exp(-kx)),$$

where the lower asymptote C is a fourth parameter to be estimated from the sample. However, this case is comparatively rare.

Following the formulae and procedure (25), we obtain

$$a' = -172{\cdot}15, \quad b = 0{\cdot}3074, \quad c' = -232{\cdot}46,$$

whence

$$\hat{y} = 50{\cdot}4, \quad k = -0{\cdot}3098, \quad f\hat{y} = -323{\cdot}4.$$

By comparison with (31) we obtain from the above parameter estimates of the exponential law, the corresponding estimates for the logistic. These are given below, alongside those obtained by Rhodes, who used a method which is computationally more cumbersome as it requires both the reciprocal, as well as a logarithmic transformation.

Example 4. U.S. Census data. Comparison of estimates of logistic parameters obtained by Rhodes with those obtained by internal regression

Parameter	Internal regression	Rhodes' (1940) method
$10{,}000/\hat{y}$	198	199
$-k$	0·3098	0·3127
$-\hat{y}f$	323	328

The agreement is extremely close, which is to be expected as the logistic curve fits the data well. In general, the theoretical properties of Rhodes's method are difficult to assess as it consists of two fitting procedures. The first (which is akin to a lag-1 autoregressive scheme) provides the estimates of k and A; these are then used to transform to linearity (as in (30)), so that the estimate of B/A is obtained as a mean of the log transforms.

(2·3) *The internal least square fit of the Makeham-Gompertz curve*

This curve is of fundamental importance in Life Table work. Its general form is

$$z = a\exp(-be^{-kx}). \tag{32}$$

Using the transformation $y = \log z$, we obtain

$$y = \log a - be^{-kx}, \tag{33}$$

which is of the form (1) or (9), so that its parameters can be estimated by the same method. We intend to deal with this application more fully elsewhere.

(2·4) *The special case of the exponential law with known asymptote (Markoff chain)*

A particular case of the first order linear difference equation is the case $a = 0$. In this case, (8) is equivalent to

$$y_{i+1} - y_i = -\tfrac{1}{2}b(y_{i+1} + y_i), \tag{34}$$

which is the systematic part of the Markoff chain equation

$$y_{i+1} = \mu y_i + \nu_i, \tag{35}$$

in which ν_i represents a random disturbance. In analogy to the general case, we have as the solution of (34)

$$y_i = y^* e^{kx_i}, \tag{36}$$

where

$$k = -2\tanh\tfrac{1}{2}b, \quad y^* = \hat{y}f. \tag{37}$$

Dealing with the case $n = \text{odd} = 2m+1$, retaining the definitions (21), we find by summation of (34)

$$\eta_j = -bY_j + w, \tag{38}$$

where w is a constant of integration. Obviously (34) and (38) are, again, equivalent equations. The least square solutions for b and w in (38) are then given by

$$\Sigma\eta_i Y_i = -b\Sigma Y_i^2 + w\Sigma Y_i, \quad \Sigma\eta_i = -b\Sigma Y_i + wn = 0. \tag{39}$$

From the second equation it follows that $w = bn^{-1}\Sigma Y_i$ and hence, from the first equation, that

$$-b = \Sigma Y_i \eta_i / \Sigma(Y_i - \bar{Y})^2. \tag{40}$$

By partial summation we may transform (40) into

$$-b = n^{-1}\Sigma\eta_i\xi_i\Sigma y_i/\Sigma(Y_i - \bar{Y})^2,$$

which is our least square estimate of b. It is proportional to the mean of the y and to the regression coefficient y on x, whilst it is inversely proportional to the sum of squares of the progressive totals Y_i of the y_i. The computational procedure is simplified as $\Sigma\eta_i\xi_i^2$ is no longer required.

3. Large sample theory of the internal least square estimators

(3·1) *The relation between internal least square and ordinary least square estimates*

In this section we compare the basic assumption on which the internal least square method is based with that made in the ordinary least square hypothesis. We confine ourselves to the special case of an exponential law with x axis as asymptote, as in (36). The general case follows on the same lines. If we denote the residuals of (36) by ϵ_i, viz.

$$\epsilon_i = y_i - y^* e^{kx_i}, \tag{41}$$

then the assumption on which the classical least square fit is based, is that the ϵ_i are independent and have the same variance. The roots y^* and k of, $\Sigma\epsilon_i^2 =$ minimum, are then maximum likelihood estimates. If on the other hand we assume that the residuals of (39), viz.

$$\zeta_i = \eta_i + b(Y_i - \bar{Y}), \tag{42}$$

are independent and have the same variance, then the internal least square estimators have the maximum likelihood property. Yet another hypothesis can be investigated:

Analogous to Bartlett's (1946) treatment of stationary time series in correlogram work, we may consider the residuals

$$\theta_i = (y_{i+1} - y_i) + \tfrac{1}{2}b(y_{i+1} + y_i) \tag{43}$$

in the difference equation (34). The assumption of independent deviates, θ_i, which is the analogue to Bartlett's starting point, would lead in the present case to the estimator

$$-\tfrac{1}{2}b = (y_m^2 - y_{-m}^2)/\Sigma(y_{i+1} + y_i)^2. \tag{44}$$

This is unsuitable, as its numerator depends on the first and last observation only. It is interesting to note the relation between the ϵ_i and ζ_i of (41) and (42). We find:

$$\zeta_i = \epsilon_i + b(E_i - \bar{E}), \tag{45}$$

where E_i and $\bar{E}$ are formed from the ϵ_i on the lines of (21). Since b is of the order n^{-1} and therefore small,† it follows that the second term in (45) is small compared with the first, particularly near the centre of the range $j = 0$. Thus the assumption of independent ζ_i does not differ seriously from that of the classical least square method. It is unlikely, therefore, that many situations will arise in which the assumption of independent ζ_i can be disproved

† It must be remembered that Y_m is a total of $\frac{1}{2}n$ values of y_i and hence, from (38), $b \sim 2\bar{y}/n\bar{y}$.

whilst the hypothesis of independent ϵ_i can be accepted. Nevertheless, we give in the next section the loss in efficiency resulting from the use of the internal least-square-fit under the more usual hypothesis that the ϵ_i are independent deviates with equal variance.

(3·2) *The efficiency of the internal least square estimator*

In this section we confine ourselves to the case $a = 0$, i.e. the exponential law with x axis as asymptote. The treatment of the general first order internal regression is on similar lines but algebraically more tedious. Further, in order to simplify the formulae we approximate exponential sums of the type $h \sum_{i=-m}^{+m} e^{kx_i} f(x_i)$ with $x_{i+1} - x_i = h$ and $n = 2m+1$, by the integrals

$$\int_{-\frac{1}{2}X}^{+\frac{1}{2}X} e^{kx} f(x)\, dx,$$

with $X = nh$. The error thereby committed is small, provided n is moderate or large. The exact summations could be carried out, but are more tedious.

We first derive the transcendental equations for the ordinary least square estimators and derive their variances from the general maximum likelihood formulae. We then derive an approximate large sample formula for the variance of the internal least square estimator b, and finally by comparison with the former, its efficiency.

(3·21) *Maximum likelihood results*

Let the range of the independent variate be $-\frac{1}{2}X \leqslant x \leqslant \frac{1}{2}X$; $X = nh$, and let us try to fit the expression

$$y(x) = y^* e^{kx} \tag{46}$$

to the observed series $y(x_i)$ where $x_i = ih$, $h = X/n$. Then, by ordinary least square or maximum likelihood procedure, we minimize

$$L = X^{-1} \int_{-\frac{1}{2}X}^{+\frac{1}{2}X} (y - y^* e^{kx})^2\, dx, \tag{47}$$

resulting in the nonlinear system of equations

$$\left.\begin{aligned} \frac{\partial L}{\partial k} &= X^{-1} \int_{-\frac{1}{2}X}^{+\frac{1}{2}X} (y - y^* e^{kx})\, e^{kx} \times y^*\, dx = 0, \\ \frac{\partial L}{\partial y^*} &= X^{-1} \int_{-\frac{1}{2}X}^{+\frac{1}{2}X} (y - y^* e^{kx})\, e^{kx}\, dx = 0. \end{aligned}\right\} \tag{48}$$

In order to obtain the variance of the maximum likelihood estimator of k ($\hat{k}$ say) we must form the Hessian

$$\Delta = \begin{vmatrix} E\left(\dfrac{\partial^2 L}{\partial k^2}\right) & E\left(\dfrac{\partial^2 L}{\partial k\, \partial y^*}\right) \\ E\left(\dfrac{\partial^2 L}{\partial k\, \partial y^*}\right) & E\left(\dfrac{\partial^2 L}{\partial y^{*2}}\right) \end{vmatrix}. \tag{49}$$

We obtain

$$\Delta = \begin{vmatrix} \frac{1}{8} X^{-1} y^2 k^{-3} \{(2+q^2)(e^q - e^{-q}) - 2q(e^q + e^{-q})\}, & \frac{1}{4} X^{-1} y k^{-2} \{q(e^q + e^{-q}) - (e^q - e^{-q})\} \\ \frac{1}{4} X^{-1} y k^{-2} \{q(e^q + e^{-q}) - (e^q - e^{-q})\}, & \frac{1}{2} X^{-1} k^{-1} (e^q - e^{-q}) \end{vmatrix}, \tag{50}$$

where $q = kX$ and k is the true parameter. Accordingly we find for the variance of $\hat{k}$ (which is given by the ratio of first minor to n times the determinant)

$$\text{Variance } \hat{k} = \sigma_\epsilon^2 8 y^{*-2} q^3 n^{-1} X^{-2} (e^q - e^{-q}) / (4q^2 + (e^q - e^{-q})^2), \tag{51}$$

which can be reduced to

$$\text{Variance } \hat{k} = \sigma_\epsilon^2 8y^{*-2}q^3 n^{-1}X^{-2}\sinh q/(\cosh 2q - 2q^2 - 1). \tag{52}$$

Letting $k \to 0$, but keeping X fixed (i.e. letting $q \to 0$) we have

$$\text{Variance } \hat{k} \to \sigma_\epsilon^2 12y^{*-2}n^{-1}X^{-2}. \tag{53}$$

As an independent check on (53) we remember that for $q \to 0$ with y^* large and fixed, the exponential law (46) will be approximately represented by the line

$$y(x) = y^* + y^*kx. \tag{54}$$

The maximum likelihood estimator of y^*k, that is the ordinary linear regression coefficient, has a variance $\sigma_\epsilon^2/\Sigma(x_i - \bar{x})^2$ which, for the present sample of x_i equidistantly spread over the interval of length X, tends to $\sigma_\epsilon^2 12/(X^2n)$, thereby confirming (53).

(3·22) *Internal least square results*

We now turn to the internal least square estimator b and study its random sampling distribution under the hypothesis that the observed values of y_i have expectations $y^*\exp(kx_i)$, from which they differ by independent random deviates ϵ_i having equal variances viz.

$$y(x_i) = y^*\exp(kx_i) + \epsilon_i. \tag{55}$$

We again use the approximation $Xn^{-1}\Sigma f(x_i) \simeq \int_{-\frac{1}{2}X}^{+\frac{1}{2}X} f(x)\,dx$.

Using elementary and partial integration and putting $kX = q$, we reach the following results:

$$\left.\begin{aligned} \text{Total of } y &= \int_{-\frac{1}{2}X}^{+\frac{1}{2}X} y(x)\,dx = \frac{y^*}{k}(e^{\frac{1}{2}q} - e^{-\frac{1}{2}q}) + \int_{-\frac{1}{2}X}^{+\frac{1}{2}X}\epsilon\,dx,\\ \bar{y} &= X^{-1}\text{ total of } y = y^*q^{-1}(e^{\frac{1}{2}q} - e^{-\frac{1}{2}q}) + \bar{\epsilon};\end{aligned}\right\} \tag{56}$$

$$\left.\begin{aligned} \eta &= y - \bar{y} = y^*e^{kx} + k\gamma + \epsilon - \bar{\epsilon},\\ \text{where}\quad \gamma &= -y^*(qk)^{-1}(e^{\frac{1}{2}q} - e^{-\frac{1}{2}q});\end{aligned}\right\} \tag{57}$$

$$Y(x) = \int_0^x y(x)\,dx = y^*k^{-1}e^{kx} + \int_0^x \epsilon\,dx; \tag{58}$$

$$Y - \bar{Y} = y^*k^{-1}e^{kx} + \gamma + \int_0^x \epsilon\,dx - X^{-1}\int_{-\frac{1}{2}X}^{+\frac{1}{2}X}\int_0^x \epsilon\,d\xi\,dx; \tag{59}$$

$$\int_{-\frac{1}{2}X}^{+\frac{1}{2}X}(Y - \bar{Y})^2\,dx = A + 2B\int_{-\frac{1}{2}X}^{+\frac{1}{2}X}\epsilon\,dx - 2\int_{-\frac{1}{2}X}^{+\frac{1}{2}X}\epsilon C(x)\,dx; \tag{60}$$

$$\left.\begin{aligned} \text{where}\quad A &= \tfrac{1}{2}y^{*2}k^{-3}(e^q - e^{-q}) - \gamma^2X,\\ B &= \tfrac{1}{2}y^*k^{-2}(e^{\frac{1}{2}q} + e^{-\frac{1}{2}q}),\\ \text{and}\quad C(x) &= y^*k^{-2}e^{kx} + \gamma x,\end{aligned}\right\} \tag{61}$$

and where the quadratic term in ϵ has been ignored. Similarly we obtain:

$$\int_{-\frac{1}{2}X}^{+\frac{1}{2}X}\eta Y\,dx = k\left(A + B\int_{-\frac{1}{2}X}^{+\frac{1}{2}X}\epsilon\,dx + \int_{-\frac{1}{2}X}^{+\frac{1}{2}X}\epsilon(-C(x) + D(x))\,dx\right), \tag{62}$$

where

$$D(x) = y^*k^{-2}e^{kx} + k^{-1}\gamma,$$

and where, again, the quadratic term in ϵ has been ignored. We now form our estimator b in accordance with (40):

$$-b = \int_{-\frac{1}{2}X}^{+\frac{1}{2}X}\eta Y\,dx \Big/ \int_{-\frac{1}{2}X}^{+\frac{1}{2}X}(Y - \bar{Y})^2\,dx, \tag{63}$$

and obtain to within linear terms in ϵ:

$$-b = k\left\{1 + A^{-1}\int_{-\frac{1}{2}X}^{+\frac{1}{2}X} \epsilon(-B + C(x) + D(x))\,dx\right\}. \tag{64}$$

We first note that to within the approximation employed the expectation of $-b$ is k, which is the first term in the expansion of $-b = 2\tanh\frac{1}{2}k$.†

Next we see that we have a representation of $-b$ in the form constant $+\int \epsilon f(x)\,dx$, i.e. a weighted sum of residuals. Remembering now that $\int \epsilon f\,dx$ is being used as an approximation to $\frac{X}{n}\Sigma f(x_i)\,\epsilon_i$ and using, therefore, the formula

$$\text{Variance}\left(\int_{-\frac{1}{2}X}^{+\frac{1}{2}X} \epsilon f(x)\,dx\right) = \sigma_\epsilon^2 X n^{-1}\int_{-\frac{1}{2}X}^{+\frac{1}{2}X} f^2(x)\,dx, \tag{65}$$

we reach

$$\text{Variance}\, b = \sigma_\epsilon^2 X n^{-1}\int_{-\frac{1}{2}X}^{+\frac{1}{2}X} (-B + C(x) + D(x))^2\,dx, \tag{66}$$

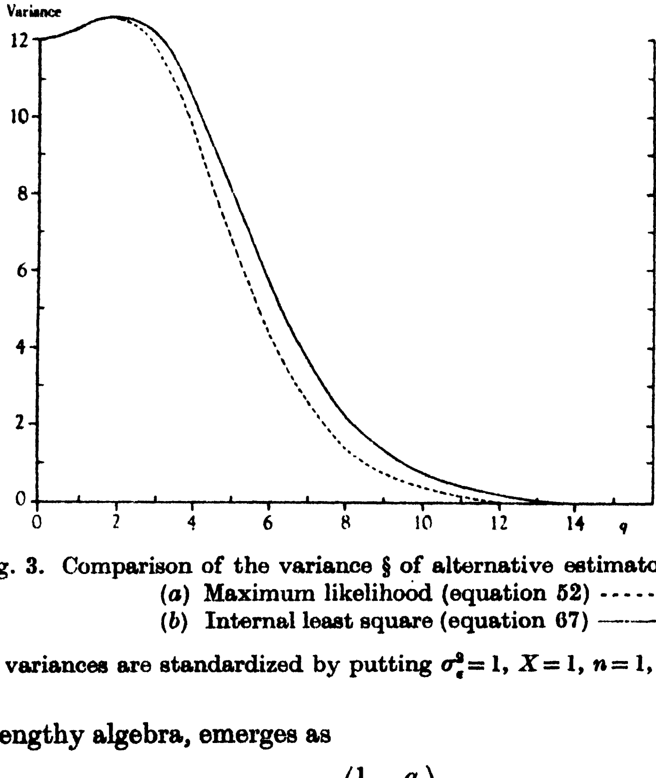


Fig. 3. Comparison of the variance § of alternative estimators:
(*a*) Maximum likelihood (equation 52) - - - - - - - -
(*b*) Internal least square (equation 67) ————

§ The variances are standardized by putting $\sigma_\epsilon^2 = 1$, $X = 1$, $n = 1$, $y^* = 1$

which, after some lengthy algebra, emerges as

$$\text{Variance}\,(b) = \sigma_\epsilon^2 \frac{4q^3}{nX^2y^{*2}}\,\frac{-1 + \left(\dfrac{1}{q} + \dfrac{q}{12}\right)\tanh\frac{1}{2}q + \frac{1}{4}q\coth\frac{1}{2}q}{\left\{1 - \dfrac{2}{q}\tanh\frac{1}{2}q\right\}^2 (e^q - e^{-q})}. \tag{67}$$

It is easy to see that for $q \to 0$, $\text{var}\, b \to 12\sigma_\epsilon^2/nX^2y^{*2}$, which is the same limit as (53) and therefore agrees with both the maximum likelihood value as well as the classical variance of the linear regression coefficient.

† For the order of discrepancy between $-b$ and k see, for instance, examples 1 to 4 of Part 2.

(3·23) *The efficiency of b*

In Fig. 3 we have plotted both the maximum likelihood variance of $\hat{k}$ as well as that of b. For given values of σ_ϵ^2, n, X and y^*, both variances are functions of $q = kX$. We have already seen that for $q \to 0$ they tend to the same limit, so that b is highly efficient for small q,† which was to be expected from the discussion given in § (3·1). As q increases both variances decrease rapidly, but the maximum likelihood variance more so and the efficiency of b drops until for large q the efficiency tends to $600/q$ %. For such large values of q (say $q \geqslant 10$), when the fitted exponential is steep and sharply bent and when $y_{-m} \geqslant 10y_m$, it is doubtful whether the basic assumption of uniform variance of the residuals ϵ_t over the whole range X is justified.

The objection that the variance of b depends on q, i.e. the parameter estimated by $-bX$, can of course be overcome (at least approximately) by an appropriate variate transformation of b. We shall not enter into this problem here.

REFERENCES

Bartlett, M. S. (1946). *J. Roy. Statist. Soc. Suppl.* **8**, 27.

Crowther, E. M. & Yates, F. (1941). *Empire J. Exp. Agric.* **9**, 77.

Cunningham, L. B. C. & Hynd, W. R. B. (1946). *J. Roy. Statist. Soc. Suppl.* **8**, 62.

Finney, D. J. (1947). *J. Roy. Statist. Soc. Suppl.* **9**, 46.

Kendall, M. G. (1944). *Biometrika*, **33**, 105.

Mann, H. B. & Wald, A. (1943). *Econometrica*, **11**, 173.

Rhodes, E. C. (1940). *J. Roy. Statist. Soc.* **103**, 362.

Sasuly, M. (1934). *Trend Analysis of Statistics*. Brookings Institution, Washington, D.C.

† Actually, on substituting in equations (52) and (67) it will be found that near $q = 2$ the maximum likelihood variance is very slightly in excess of that of the internal least square solution, an anomaly caused by the approximations on which (67) is based.

THE GEOMETRICAL METHOD IN THE THEORY OF SAMPLING

By DAVID FOG, *den kgl. Veterinaer- og Landbohøjskole, Copenhagen*

For the determination of exact sampling distributions a number of different methods have been employed, of which two will be mentioned here:

(1) The geometrical method, which has been applied with great success especially by R. A. Fisher (1915, 1925) and consists in the use of the terminology and results of multidimensional geometry;

(2) the analytical method, the main line of which is a change of variables and calculation of the corresponding functional determinants.

The first method appears elegant and short, whereas the second frequently leads to long calculations. On the other hand, the view is often expressed that the geometrical method arrives at its results too easily, at the cost of the accuracy usually required, and which is duly honoured by the analytical method. Probably this view has been affirmed by the circumstance that it has proved difficult to translate the geometrical methods and considerations so as to enter naturally and smoothly into the usual analytical treatment.

With regard to the criticism that the geometrical method is less accurate, it should be observed that the justification depends entirely on the way in which the multidimensional geometry is built up. The construction of the multidimensional geometry may be worked out on a purely numerical basis, without introducing a single geometrical axiom, but in such a manner that the meaning of the geometrical terms as well as the contents of the geometrical formulae are of an entirely numerical nature. The system then rests on the same basis as the usual mathematical analysis and is therefore just as precise as the latter.

This subject will not be discussed further here, but we shall show in the present paper how in a number of cases the geometrical method may, with complete preservation of its simplicity, be translated into analytical form, so that, without any knowledge of multidimensional geometry, advantage may be taken of the simple methods to which it has given rise.

Before dealing with the examples we shall recall the concepts of volume and surface-content of an n-dimensional sphere. Such a sphere lies in an n-dimensional space and may be represented by the equation

$$x_1^2+x_2^2+\ldots+x_n^2 = a^2; \tag{1}$$

its interior is the domain

$$x_1^2+x_2^2+\ldots+x_n^2 < a^2. \tag{2}$$

The volume of this domain is determined as the multiple integral

$$V_n(a) = \int_{\sum_1^n x_i^2<a^2} \cdots\cdots \int dx_1\,dx_2\ldots dx_n, \tag{3}$$

and it is easily shown that

$$V_n(a) = \frac{\pi^{\frac{1}{2}n}a^n}{\Gamma(\frac{1}{2}n+1)}. \tag{4}$$

In order to find the corresponding surface-content $S_n(a)$ we consider $dV_n(a)$. This denotes the volume between two concentric spheres of radii a and $a+da$; thus $dV_n(a) = S_n(a)\,da$. We therefore obtain

$$S_n(a) = \frac{dV_n(a)}{da} = \frac{2\pi^{\frac{1}{2}n}a^{n-1}}{\Gamma(\frac{1}{2}n)}. \tag{5}$$

It should be noted that the result (4), even with exclusion of all geometry, has a definite meaning as the value of the multiple integral (3). As for (5), from a purely analytical point of view, we need only regard it in the following as a definition of $S_n(a)$.

We shall now consider three examples. The first two are quite elementary, and the results of all three are well known. Each example will be treated in two ways, in a geometrical and in an analytical manner, the latter being the translation of the former into analytical language; the essential points are the possibility of this translation and its method of working out.

Example 1. Let $x_1, x_2, \ldots, x_n$ be n mutually independent variables, all normal $(0,\sigma)$, i.e. having the frequency function

$$p\{x\} = \frac{1}{\sqrt{(2\pi)}\,\sigma}\exp\left[-\frac{x^2}{2\sigma^2}\right]. \tag{6}$$

The frequency function of the set $X = (x_1, x_2, \ldots, x_n)$ is then

$$\kappa = p\{X\} = \frac{1}{(\sqrt{(2\pi)}\,\sigma)^n}\exp\left[-\frac{\sum\limits_1^n x_i^2}{2\sigma^2}\right]. \tag{7}$$

Putting

$$q = \sqrt{\left(\sum_1^n x_i^2\right)}, \tag{8}$$

the formula (7) may also be written

$$\kappa = \frac{1}{(\sqrt{(2\pi)}\,\sigma)^n}\exp\left[-\frac{q^2}{2\sigma^2}\right]. \tag{9}$$

Assuming X to be a point in an n-dimensional space and $\kappa = p\{X\}$ to be the density at this point, the probability of X being in a domain ω is equal to the mass of ω and is given by the multiple integral

$$\int \ldots \int \kappa\, dX, \tag{10}$$

taken throughout ω, where $dX = dx_1\,dx_2 \ldots dx_n$ is an element of volume.

This in particular comprises the determination of the distribution of an arbitrary statistic

$$y = \phi(x_1, x_2, \ldots, x_n).$$

Designating its distribution function as $P\{y\}$, (10) gives $dP\{y\}$, when for ω we put the domain

$$y < \phi(x_1, x_2, \ldots, x_n) < y + dy.$$

We shall now find the distribution of the statistic q, introduced in (8). This variable may be interpreted as the distance from the origin O to the point X. Hence we use as element of volume the region between two n-dimensional spheres of common centre O and radii q and $q+dq$ and obtain

$$dP\{q\} = \kappa\, dV_n(q) = \kappa S_n(q)\, dq.$$

Inserting the values of κ and $S_n(q)$ from (9) and (5) we get

$$dP\{q\} = \frac{1}{2^{\frac{1}{2}(n-2)}\Gamma(\frac{1}{2}n)}\exp\left[-\frac{q^2}{2\sigma^2}\right]\left(\frac{q}{\sigma}\right)^{n-1}\frac{dq}{\sigma}, \tag{11}$$

and we have obtained the distribution required. For $\sigma = 1$ (11) in particular shows the distribution of χ.

Next we will prove (11) without any use of geometry, but following the exposition above as closely as possible. We then start with (9)—where q is determined by (8)—and

$$dP\{X\} = \kappa dX. \tag{12}$$

By an almost* one-to-one correspondence we pass from the variables $X = (x_1, x_2, \ldots, x_n)$ to

$$q, \quad U = (u_1, u_2, \ldots, u_{n-1}),$$

the u's being co-ordinates on the unit-sphere, the specification of which is of no importance for the proof.† Designating by Δ the jacobian $\partial(X)/\partial(q, U)$, we get

$$dP\{q, U\} = \kappa \mid \Delta \mid dq dU,$$

from which we find
$$dP\{q\} = \kappa dq \int_{\omega}\!\!\ldots\!\int \mid \Delta \mid dU, \tag{13}$$

ω being the domain of the u's.

Hence it remains only to determine the multiple integral in (13), which is done as follows:

From
$$V_n(a) = \int_{\sum_1^n x_i^2 < a^2} \!\!\ldots\ldots \int dX$$

we obtain by means of the substitution used above

$$V_n(a) = \int\!\!\ldots\!\int \mid \Delta \mid dq dU = \int_0^a dq \int_{\omega}\!\!\ldots\!\int \mid \Delta \mid dU,$$

thus by differentiation
$$\frac{dV_n(q)}{dq} = \int_{\omega}\!\!\ldots\!\int \mid \Delta \mid dU.$$

Hence the integral desired is equal to $S_n(q)$, and (13) becomes

$$dP\{q\} = \kappa S_n(q)\, dq,$$

which, just as before, leads to (11).

Note. The assumption that the true mean of x is 0 is of little importance. For if this mean is equal to $\mu \neq 0$, we have only to consider, instead of x, the deviation $x - \mu$ from μ. We then find that

$$q = \sqrt{\left(\sum_1^n (x_i - \mu)^2\right)}$$

has the distribution (11). A similar modification will apply to the following examples.

Example 2. As before, let $X = (x_1, x_2, \ldots, x_n)$ be distributed so that (6) and (7) are valid. From the x's we form the linear combination

$$z = \sum_1^n a_i x_i, \tag{14}$$

where for the sake of convenience we may assume that

$$\sum_1^n a_i^2 = 1. \tag{15}$$

* The exceptional points forming a domain of volume 0.

† A suitable transformation is for instance

$$\begin{aligned}
x_1 &= q \cos u_1 && (q > 0),\\
x_2 &= q \sin u_1 \cos u_2 && (0 < u_i < \pi,\ i = 1, 2, \ldots, n-2),\\
&\ldots\\
x_{n-1} &= q \sin u_1 \sin u_2 \ldots \sin u_{n-2} \cos u_{n-1} && (0 < u_{n-1} < 2\pi),\\
x_n &= q \sin u_1 \sin u_2 \ldots \sin u_{n-2} \sin u_{n-1}.
\end{aligned}$$

In Fig. 1, α represents the hyperplane $\sum_1^n a_i x_i = 0$, passing through O; the line l is perpendicular to α at O. The point X is projected on l and α as N and Q; the quadrilateral $ONXQ$ is a rectangle. The co-ordinates of N and Q are easily found to be

$$N = (a_1 z, \ldots, a_n z), \quad Q = (x_1 - a_1 z, \ldots, x_n - a_n z).$$

Further $ON = z$, and for OQ, denoted by q_1, we find

$$q_1^2 = \sum_1^n (x_i - a_i z)^2. \tag{16}$$

From the right-angled triangle OQX, where $OX^2 = \sum_1^n x_i^2$, we have

$$\sum_1^n x_i^2 = z^2 + q_1^2, \tag{17}$$

which may also be easily verified directly by calculation.

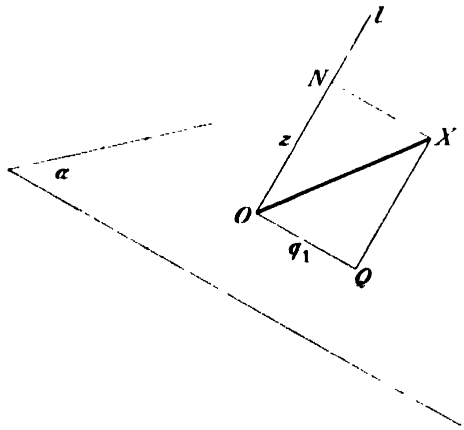


Fig. 1.

By means of (17), we may write (7) as

$$\kappa = \frac{1}{(\sqrt{(2\pi)}\,\sigma)^n} \exp\left[-\frac{z^2 + q_1^2}{2\sigma^2}\right]. \tag{18}$$

We will now find the joint distribution of z and q_1. As element of volume for X we use a domain, the projection of which on α is the region between two $(n-1)$-dimensional spheres in α, having the common centre O and the radii q_1 and $q_1 + dq_1$, and the projection of which on l is the line-element dz. We then get

$$dP\{z, q_1\} = \kappa\, dV_{n-1}(q_1)\, dz = \kappa S_{n-1}(q_1)\, dq_1\, dz. \tag{19}$$

Hence by (18) and (5)

$$dP\{z, q_1\} = \frac{1}{\sqrt{(2\pi)}\,\sigma} \exp\left[-\frac{z^2}{2\sigma^2}\right] dz \frac{1}{2^{\frac{1}{2}(n-3)}\Gamma\frac{1}{2}(n-1)} \exp\left[-\frac{q_1^2}{2\sigma^2}\right] \left(\frac{q_1}{\sigma}\right)^{n-2} \frac{dq_1}{\sigma}. \tag{20}$$

Thus z and q_1 are independent; z is normal $(0, \sigma)$, and q_1 is distributed as q in (11), but with $n-1$ instead of n.

If in particular $a_1 = a_2 = \ldots = a_n = 1/\sqrt{n}$, we get, $\bar{x}$ being the mean $\sum_1^n x_i/n$,

$$z = \bar{x}\sqrt{n}, \quad q_1^2 = \sum_1^n (x_i - \bar{x})^2.$$

Let us now, without the aid of geometry, once more prove (20), starting from (6), (7), (14), (15) and

$$dP\{X\} = \kappa dX. \tag{21}$$

We introduce new variables z, $Y = (y_1, y_2, \ldots, y_{n-1})$ by means of an orthogonal substitution

$$z = a_1x_1 + a_2x_2 + \ldots + a_nx_n, \quad y_i = a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n \quad (i = 1, 2, \ldots, n-1), \tag{22}$$

and put

$$q_1 = \sqrt{\left(\sum_1^{n-1} y_i^2\right)}. \tag{23}$$

As (22) leaves sums of squares invariant we get

$$\sum_1^n x_i^2 = z^2 + q_1^2. \tag{24}$$

A comparison with (17) shows that the q_1's introduced by (23) and (16) are identical.

The jacobian corresponding to (22) being ± 1 we get from (21)

$$dP\{z, Y\} = \kappa dz dY, \tag{25}$$

where as before κ is determined by (18).

By a transformation analogous to the one on p. 48, but with $n-1$ instead of n, we may instead of Y introduce new variables $(q_1, u_1, u_2, \ldots, u_{n-2})$. After integration relative to the u's we have

$$dP\{z, q_1\} = \kappa dz S_{n-1}(q_1)\, dq_1,$$

which is equivalent to (19) and immediately leads to (20).

Example 3. As a final example we consider the normal correlation between k variables. In the geometrical treatment we use a method employed by Wishart (1928) for the derivation of the distribution named after him, with certain simplifications, however, which will facilitate the transition to the analytical exposition. Furthermore, for convenience, we assume in the following $k = 3$; the generalization from $k = 3$ to an arbitrary value should cause no difficulties, the addition of one more variable merely introducing another step of reduction.

Let x, y, z denote three normally correlated variables. Assuming that the three true means are equal to 0, the frequency function may be written

$$p\{x, y, z\} = \frac{\sqrt{A}}{(2\pi)^{\frac{3}{2}}} \exp\left[-\tfrac{1}{2}(a_{11}x^2 + 2a_{12}xy + \ldots + a_{33}z^2)\right], \tag{26}$$

where A is the determinant $|a_{\mu\nu}|$. We consider n triples (x_i, y_i, z_i), mutually independent and each satisfying (26). We introduce the designations

$$l_{11} = \sum_1^n x_i^2, \quad l_{12} = \sum_1^n x_i y_i, \quad \ldots, \quad l_{33} = \sum_1^n z_i^2. \tag{27}$$

Putting furthermore in the same way as before

$$X = (x_1, x_2, \ldots, x_n), \quad Y = (y_1, y_2, \ldots, y_n), \quad Z = (z_1, z_2, \ldots, z_n), \tag{28}$$

we get

$$\kappa = p\{X, Y, Z\} = \frac{A^{\frac{1}{2}n}}{(2\pi)^{\frac{3}{2}n}} \exp\left[-\tfrac{1}{2}(a_{11}l_{11} + 2a_{12}l_{12} + \ldots + a_{33}l_{33})\right]. \tag{29}$$

In the following our aim is to find the distribution of the quantities $l_{11}, l_{12}, \ldots, l_{33}$. κ in (29) denotes the density at the point

$$M = (x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n, z_1, z_2, \ldots, z_n) \tag{30}$$

in a $3n$-dimensional space; the origin is denoted O. Within this space we consider three n-dimensional subspaces, corresponding to the first n, the intermediate n, and the last n co-ordinates. The projections of M on these three subspaces are the points X, Y, and Z in (28)

in the sense that for each point only the n co-ordinates belonging to the corresponding subspace are indicated (the other $2n$ being zero).

For the time being it will be convenient to imagine that Y and Z, with the co-ordinates shown in (28), are placed in the same n-dimensional subspace as X. Thereby we are, *inter alia*, enabled to give a simple geometrical interpretation of all the l's. Thus l_{11} means the square of the distance OX, while l_{12} means the product $OX.OY.\cos\phi_{12}$, where ϕ_{12} denotes the angle XOY, etc.

In the n-dimensional space just mentioned we introduce a new co-ordinate system with the same origin O and in close relation to the points X, Y, and Z (Fig. 2): the x_1-axis is chosen along OX, the x_2-axis perpendicular to OX in the plane OXY, and the x_3-axis perpendicular to the plane OXY in the three-dimensional space $OXYZ$. The remaining $n-3$ axes may be chosen arbitrarily (subject only to the conditions of orthogonality).

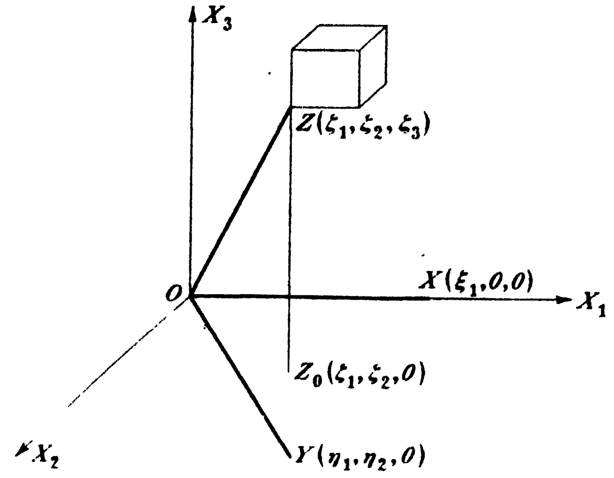


Fig. 2.

The new co-ordinates of the points X, Y and Z are denoted as in Fig. 2 (where only the first three are shown, the remaining being zero). Further we may assume that the change of co-ordinates is carried out in such a manner that $\xi_1 > 0$, $\eta_2 > 0$ and $\zeta_3 > 0$.

The l's being invariant by change of co-ordinates and hence expressed in the same manner by the new and the old co-ordinates, we have

$$\left.\begin{aligned} l_{11} &= \xi_1^2, \quad l_{12} = \xi_1\eta_1, \quad && l_{13} = \xi_1\zeta_1, \\ & \quad\; l_{22} = \eta_1^2+\eta_2^2, && l_{23} = \eta_1\zeta_1+\eta_2\zeta_2, \\ & && l_{33} = \zeta_1^2+\zeta_2^2+\zeta_3^2. \end{aligned}\right\} \tag{31}$$

In relation to any set of points X, Y, Z in the n-dimensional space we may introduce a new co-ordinate system in the above manner and determine the six characteristic quantities $\xi_1, \eta_1, \eta_2, \zeta_1, \zeta_2, \zeta_3$. To begin with the distribution of these six quantities is sought.

With that in view we consider such sets of points X, Y, Z, for which the six quantities have fixed values and where, moreover, the points X and Y are fixed. The projection $Z_0 = (\zeta_1, \zeta_2, 0)$ of Z on the plane OXY will also be fixed, and the distance of Z from the plane has the constant value ζ_3. The point Z, therefore, will be able to move on the surface of an $(n-2)$-dimensional sphere of centre Z_0 and radius ζ_3, situated in an $(n-2)$-dimensional space perpendicular to the plane OXY in Z_0. To Z we attach (see Fig. 2) a small element of

volume $d\zeta_1 d\zeta_2 d\zeta_3$. When Z moves on its spherical surface this element of volume will describe an n-dimensional volume

$$S_{n-2}(\zeta_3)\,d\zeta_1 d\zeta_2 d\zeta_3. \tag{32}$$

Next we consider such sets of points X, Y, where ξ_1, η_1 and η_2 have fixed values, and where moreover X is fixed. The point Y will then be able to move on an $(n-1)$-dimensional sphere of radius η_2 and centre in the projection of Y on OX; a small element of area $d\eta_1 d\eta_2$ attached to Y will thereby describe a volume

$$S_{n-1}(\eta_2)\,d\eta_1 d\eta_2. \tag{33}$$

Finally, when X moves in such a manner that the value of ξ_1 is fixed, then a line-element $d\xi_1$ attached to X will describe an element of volume

$$S_n(\xi_1)\,d\xi_1. \tag{34}$$

We now return to the point M in (30), the projections of which on three mutually orthogonal, n-dimensional subspaces are X, Y and Z in their original positions. To a given set of values $\xi_1, \eta_1, \eta_2, \zeta_1, \zeta_2, \zeta_3$ with corresponding differentials $d\xi_1, d\eta_1, d\eta_2, d\zeta_1, d\zeta_2, d\zeta_3$ we obtain in the $3n$-dimensional space an element of volume equal to the product of the three elements (32), (33) and (34), i.e.

$$S_n(\xi_1)\,S_{n-1}(\eta_2)\,S_{n-2}(\zeta_3)\,d\xi_1 d\eta_1 d\eta_2 d\zeta_1 d\zeta_2 d\zeta_3. \tag{35}$$

Hence we have

$$dP\{\xi_1, \eta_1, \eta_2, \zeta_1, \zeta_2, \zeta_3\} = \kappa S_n(\xi_1)\,S_{n-1}(\eta_2)\,S_{n-2}(\zeta_3)\,d\xi_1 d\eta_1 d\eta_2 d\zeta_1 d\zeta_2 d\zeta_3, \tag{36}$$

where κ is given by (29). Thus we have found the required distribution of $\xi_1, \eta_1, \eta_2, \zeta_1, \zeta_2, \zeta_3$, as by means of the formulae (31) these six quantities may be inserted in κ.

Substituting the values of the spherical surfaces in (36) we get

$$dP\{\xi_1, \eta_1, \ldots, \zeta_3\} = \kappa \frac{2\pi^{\frac{1}{2}n}\xi_1^{n-1}}{\Gamma(\frac{1}{2}n)} \frac{2\pi^{\frac{1}{2}(n-1)}\eta_2^{n-2}}{\Gamma\frac{1}{2}(n-1)} \frac{2\pi^{\frac{1}{2}(n-2)}\zeta_3^{n-3}}{\Gamma\frac{1}{2}(n-2)}\,d\xi_1 d\eta_1 \ldots d\zeta_3,$$

thus

$$dP\{\xi_1, \eta_1, \ldots, \zeta_3\} = \kappa \frac{8\pi^{\frac{1}{2}(3n-3)}}{\Gamma(\frac{1}{2}n)\,\Gamma\frac{1}{2}(n-1)\,\Gamma\frac{1}{2}(n-2)}\,\xi_1^{n-1}\eta_2^{n-2}\zeta_3^{n-3}\,d\xi_1 d\eta_1 \ldots d\zeta_3. \tag{37}$$

We shall now find the distribution of the l's introduced in (27). The jacobian of the l's in relation to $\xi_1, \eta_1, \ldots, \zeta_3$ may be split up into a product of three and becomes

$$\frac{dl_{11}}{d\xi_1}\frac{\partial(l_{12}, l_{22})}{\partial(\eta_1, \eta_2)}\frac{\partial(l_{13}, l_{23}, l_{33})}{\partial(\zeta_1, \zeta_2, \zeta_3)} = 2\xi_1 \begin{vmatrix} \xi_1 & 0 \\ 2\eta_1 & 2\eta_2 \end{vmatrix} \begin{vmatrix} \xi_1 & 0 & 0 \\ \eta_1 & \eta_2 & 0 \\ 2\zeta_1 & 2\zeta_2 & 2\zeta_3 \end{vmatrix} = 8\xi_1^3\eta_2^2\zeta_3.$$

Hence we get

$$dP\{l_{11}, l_{12}, \ldots, l_{33}\} = \kappa \frac{\pi^{\frac{1}{2}(3n-3)}}{\Gamma(\frac{1}{2}n)\,\Gamma\frac{1}{2}(n-1)\,\Gamma\frac{1}{2}(n-2)}\,(\xi_1\eta_2\zeta_3)^{n-4}\,dl_{11}dl_{12}\ldots dl_{33}. \tag{38}$$

Moreover, by means of the rule of multiplying determinants we have

$$\begin{vmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{vmatrix} = \begin{vmatrix} \xi_1 & 0 & 0 \\ \eta_1 & \eta_2 & 0 \\ \zeta_1 & \zeta_2 & \zeta_3 \end{vmatrix}^2 = (\xi_1\eta_2\zeta_3)^2,$$

whence (38) becomes

$$dP\{l_{11}, l_{12}, \ldots, l_{33}\} = \kappa \frac{\pi^{\frac{1}{2}(3n-3)}}{\Gamma(\frac{1}{2}n)\,\Gamma\frac{1}{2}(n-1)\,\Gamma\frac{1}{2}(n-2)} \begin{vmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{vmatrix}^{\frac{1}{2}(n-4)} dl_{11}dl_{12}\ldots dl_{33}. \tag{39}$$

This is the required distribution of the l's, as κ is expressed by the l's in (29).

We shall now repeat the proof in a manner which will not involve any geometrical aids. To do this, we proceed once more from (26) and introduce the l's by (27). Furthermore, we introduce X, Y and Z by (28); then

$$dP\{X, Y, Z\} = \kappa dX dY dZ, \tag{40}$$

where κ is determined by (29). We may for the sake of convenience continue to use the words 'point' and 'co-ordinate'.

We consider an orthogonal transformation

$$\left.\begin{aligned} t_1' &= \frac{1}{\xi_1}(x_1 t_1 + \dots + x_n t_n) \quad & \xi_1 &= \sqrt{\left(\sum_1^n x_i^2\right)}, \\ t_i' &= \alpha_{i1} t_1 + \dots + \alpha_{in} t_n \quad & i &= 2, 3, \dots, n, \end{aligned}\right\} \tag{I}$$

carrying $(t_1, t_2, \dots, t_n)$ into $(t_1', t_2', \dots, t_n')$. Applied to the point X it gives

$$X' = (\xi_1, 0, 0, \dots, 0),$$

all the co-ordinates of X' except the first being equal to zero owing to the conditions of orthogonality

$$\alpha_{i1} x_1 + \alpha_{i2} x_2 + \dots + \alpha_{in} x_n = 0 \quad (i = 2, 3, \dots, n).$$

By (I) the points Y and Z are transformed into

$$Y' = (\eta_1, y_2', \dots, y_n'), \quad Z' = (\zeta_1, z_2', \dots, z_n'),$$

where the co-ordinates need not be calculated. The special symbols η_1 and ζ_1 are due to the fact that these co-ordinates will not change by the two following transformations.

Next we use a new orthogonal transformation

$$\left.\begin{aligned} t_1'' &= t_1', \\ t_2'' &= \frac{1}{\eta_2}(y_2' t_2' + \dots + y_n' t_n') \quad & \eta_2 &= \sqrt{\left(\sum_2^n y_i'^2\right)}, \\ t_i'' &= \beta_{i2} t_2' + \dots + \beta_{in} t_n' \quad & i &= 3, 4, \dots, n. \end{aligned}\right\} \tag{II}$$

This transformation does not change X', but carries Y' and Z' into

$$Y'' = (\eta_1, \eta_2, 0, \dots, 0), \quad Z'' = (\zeta_1, \zeta_2, z_3'', \dots, z_n'').$$

η_1 and ζ_1 remain unchanged as noted above. The last $n-1$ co-ordinates of Z'' need not be calculated. The special symbol ζ_2 is due to the fact that this quantity is unchanged by the following transformation.

Finally, we use a third orthogonal transformation

$$\left.\begin{aligned} t_1''' &= t_1'', \\ t_2''' &= t_2'', \\ t_3''' &= \frac{1}{\zeta_3}(z_3'' t_3'' + \dots + z_n'' t_n'') \quad & \zeta_3 &= \sqrt{\left(\sum_3^n z_i''^2\right)}, \\ t_i''' &= \gamma_{i3} t_3'' + \dots + \gamma_{in} t_n'' \quad & i &= 4, 5, \dots, n. \end{aligned}\right\} \tag{III}$$

This transformation changes neither X' nor Y'', but carries Z'' into

$$Z''' = (\zeta_1, \zeta_2, \zeta_3, 0, \dots, 0).$$

Thus we have once more introduced the quantities ξ_1, η_1, η_2, ζ_1, ζ_2, ζ_3, known from the geometrical treatment, and as the l's are invariant by orthogonal transformations we again

We now return to (40) and apply some transformations of variables:

(1) For Y and Z we introduce Y' and Z', respectively, whereas X is left unchanged. As the coefficients of (I) depend only on X, the jacobians are equal to ± 1, and (40) is transformed into

$$dP\{X, \eta_1, Y'_1, \zeta_1, Z'_1\} = \kappa\, dX\, d\eta_1\, dY'_1\, d\zeta_1\, dZ'_1, \tag{41}$$

where $Y'_1 = (y'_2, \ldots, y'_n)$ and $Z'_1 = (z'_2, \ldots, z'_n)$.

(2) For Z'_1 we introduce the last $n-1$ co-ordinates from Z'', while X, η_1, and Y'_1 as well as ζ_1 are unchanged. As the coefficients of (II) depend only on Y'_1 the jacobian as above is equal to ± 1, and (41) becomes

$$dP\{X, \eta_1, Y'_1, \zeta_1, \zeta_2, Z''_2\} = \kappa\, dX\, d\eta_1\, dY'_1\, d\zeta_1\, d\zeta_2\, dZ''_2, \tag{42}$$

where $Z''_2 = (z''_3, \ldots, z''_n)$.

According to (31) κ may be expressed by the variables $\xi_1, \eta_1, \eta_2, \zeta_1, \zeta_2, \zeta_3$ only. Hence, as in example 1, instead of X we may introduce $\xi_1 = \sqrt{\left(\sum_1^n x_i^2\right)}$ together with $U = (u_1, u_2, \ldots, u_{n-1})$, for Y'_1 similarly $\eta_2 = \sqrt{\left(\sum_2^n y_i'^2\right)}$ together with $V = (v_1, v_2, \ldots, v_{n-2})$ and finally for Z''_2, $\zeta_3 = \sqrt{\left(\sum_3^n z_i''^2\right)}$ and $W = (w_1, w_2, \ldots, w_{n-3})$ as new variables. Eliminating U, V and W by integration we get from (42)

$$dP\{\xi_1, \eta_1, \eta_2, \zeta_1; \zeta_2, \zeta_3\} = \kappa S_n(\xi_1)\, S_{n-1}(\eta_2)\, S_{n-2}(\zeta_3)\, d\xi_1\, d\eta_1\, d\eta_2\, d\zeta_1\, d\zeta_2\, d\zeta_3.$$

Thus we have once more proved (36) and may as before obtain (37) as well as (39).

Just as we have seen in example 2 that $\sum_1^n (x_i - \bar{x})^2$ is distributed as $\sum_1^n x_i^2$, but with $n-1$ instead of n, it may be shown that the variables

$$\bar{l}_{11} = \sum_1^n (x_i - \bar{x})^2, \quad \bar{l}_{12} = \sum_1^n (x_i - \bar{x})(y_i - \bar{y}), \quad \ldots, \quad \bar{l}_{33} = \sum_1^n (z_i - \bar{z})^2$$

are distributed as the l's in (27), but with $n-1$ instead of n, $\bar{x}$, $\bar{y}$ and $\bar{z}$ being the means of x_i, y_i and z_i respectively. The proof is omitted here.

Finally, it should be noted that the constants of the distributions dealt with in the above all enter automatically and need not be determined by a special process. The common source of all these constants is formula (4), giving the volume of the n-dimensional sphere.

REFERENCES

Fisher, R. A. (1915). *Biometrika*, **10**, 507.
Fisher, R. A. (1925). *Metron*, **5**, 90.
Wishart, J. (1928). *Biometrika*, **20 A**, 32.

PROOFS OF THE DISTRIBUTION LAW OF THE SECOND ORDER MOMENT STATISTICS

By JOHN WISHART, *School of Agriculture, Cambridge*

The foregoing paper by Prof. Fog (1948) calls to mind the different occasions on which proofs have been derived, by various methods, of the result first given twenty years ago by the present author (Wishart, 1928). It may be useful at this stage to catalogue these proofs, so far as known to the writer, and to comment on the contrasting methods employed.

The method first used was a direct extension of the geometrical method of approach used by R. A. Fisher (1915), and was worked out in full for three variates, then extended to the general case by the use of quadratic co-ordinates. At the end of 1933 Prof. Mahalanobis sent the author a somewhat fuller proof on the same lines, which was published some years later (Mahalanobis, Bose & Roy, 1937) as part of a long study of the normalization of statistical variates and the use of rectangular co-ordinates in the theory of sampling distributions. A proof on entirely different lines had been published before the communication referred to from Prof. Malahanobis was received (Wishart & Bartlett, 1933). This depended upon working out the characteristic function of the distribution and then using a generalization of the Fourier integral theorem to express the result as a multiple integral, detailed consideration to which was given by Ingham (1933). Special cases of this method, and of the resulting integrals, had previously been considered for two variables by Romanovsky (1925).

Part of the difficulty, as in all similar problems in the theory of sampling distributions, resides in the fact that the practical problem requires the moment statistics to be calculated from the sample means. This general feature, which is allowed for by writing $n-1$ for n, where n is the size of the sample, may be separated from the fundamental distribution problem, which is that of the simultaneous variation of the $\frac{1}{2}k(k+1)$ sums of squares and products of k variate values $x_i(i=1, 2, \ldots k)$, which may be taken to follow a multivariate normal distribution with zero means. This was recognized by P. L. Hsu (1939), who obtained a proof by the method of mathematical induction which has the merit of being short. Madow (1938) deduced the distribution by generalizing from Hotelling's joint distribution of sample correlation coefficients. Of the text-book proofs those of Wilks (1943) and Kendall (1946) are geometrical, and that of Cramér (1946) determines the characteristic function. Cramér also refers to a paper by Simonsen (1944, 1945).

The distribution is a direct generalization of the χ^2 distribution to the case of vectors with a number k (>1) of components. Reference should also be made to Bartlett (1933) who deduced the corresponding partial distribution when a number of variates, about whose distribution nothing need be assumed, are eliminated, and to Anderson & Girshick (1944; Anderson, 1946) who deduced cases of the corresponding non-central distribution analogous to the similar problem in the case of χ^2. Recently Elfving (1947) has used matrix methods to deduce Bartlett's decomposition theorem (Bartlett, 1933) and states that the general distribution can be proved in this way, although a formal proof is not given.

Prof. Fog's proof, which provides an interesting parallelism between the geometrical and analytical approaches to the problem, is only given for the case of three variates, but in contradistinction to former proofs his steps would seem to be such as to justify his statement that the generalization to an arbitrary value of k should cause no difficulties. It may not be out of place here to remark that the methods used by him in Examples 1 and 2 may be

used to give a geometrical interpretation to certain special orthogonal transformations which have been used in deriving tests of significance. It is true that the variables U need not be particularly specified, but it helps the elementary student if, for example, $\sum_1^n (x-\bar{x})^2$ and, in the regression problem, $\sum_1^n (y-a-b(x-\bar{x}))^2$ can be directly expressed in a convenient way as the sums of squares of $n-1$ and $n-2$ independent normal variates respectively. A recent expository paper (Wishart, 1947) deduced all the results required for the proof of the variance ratio distribution by using the orthogonal transformation

$$u_i = -\sqrt{\left(\frac{i}{i+1}\right)}(x_{i+1}-\bar{x}_i) \quad (i = 1, 2, \dots, n-1),$$
$$u_n = \sqrt{n}\,.\,\bar{x}_n,$$

where $x_1, x_2, \dots, x_n$ are normal (0, 1) and $\bar{x}_i = \sum_{r=1}^{i} (x_r)/i$. Then $\sum_1^n (x_i-\bar{x}_n)^2 = \sum_1^{n-1} (u_i^2)$. Similarly, for a regression relationship of the form $Y = bx$ we can make the transformation (Vajda 1945):

$$v_i = -\sqrt{\left(\frac{\Sigma_i}{\Sigma_{i+1}}\right)}(y_{i+1}-b_i x_{i+1}) \quad (i = 1, 2, \dots, n-1),$$
$$v_n = \sqrt{\Sigma_n}\,.\,b_n,$$

where $y_1, y_2, \dots, y_n$ are normal (0, 1), the x_i can be regarded as fixed, and

$$b_i = \sum_{r=1}^{i} (x_r y_r)/\Sigma_i, \quad \Sigma_i = \sum_{r=1}^{i} (x_r^2).$$

Then
$$\sum_1^n (y_i - b_n x_i)^2 = \sum_1^{n-1} (v_i^2).$$

In both cases (the first is a particular case of the second, in which x_i is put equal to unity for all i), the figure which applies is Fig. 1 of Prof. Fog's paper. $OQ^2 = q_1^2$ represents $\Sigma(x-\bar{x})^2$ in the one case and $\Sigma(y-bx)^2$ in the other, since the linear relations are $\Sigma(x/\sqrt{n}) = \sqrt{n}\,.\,\bar{x}$ and $\Sigma(yx/\sqrt{\Sigma_n}) = \sqrt{\Sigma_n}\,.\,b$ respectively, and this quantity is therefore the sum of the squares of $(n-1)$ perpendicular distances from Q on to $(n-1)$ hyperplanes through the origin orthogonal to α, and mutually orthogonal among themselves. Any one of an infinity of planes can be chosen for the first, and this corresponds to the statement that the variables U need not be specified. In our first illustration the plane is chosen as $x_1 = x_2$, followed by $x_1 + x_2 = 2x_3$, etc., whereas in the second illustration we start with $x_2 y_1 = x_1 y_2$ (where y_1 and y_2 are the variables, and the x's are constants) and follow up with

$$x_3(x_1y_1 + x_2y_2) = y_3(x_1^2 + x_2^2), \text{ etc.}$$

Finally, the distributions required are easily deduced from the fact that the joint probability, after transformation, breaks up into the product of independent parts.

The more general regression relationship $Y = a + b(x-\bar{x})$ is dealt with by applying the first transformation to both the x's and y's, reducing the sums of n squares and products to the sums of $(n-1)$ squares and products of new variables u and u'. To u' is now applied the second transformation (using u) to a variable v', in which i goes from 1 to $n-2$, and

$$v'_{n-1} = \sqrt{\Sigma_{n-1}}\,.\,b_{n-1},$$

where
$$\Sigma_{n-1} = \sum_{r=1}^{n-1} (u_r^2), \quad b_{n-1} = \sum_{r=1}^{n-1} (u_r u'_r)/\Sigma_{n-1}.$$

We then have
$$\sum_1^n (y-Y)^2 = \sum_1^{n-2} (v_i'^2)$$

(Wishart, 1948). This represents a particular application of Fisher's general theorem (1925).

Geometrically, the position is as represented in the diagram, which is a further development of Prof. Fog's Fig. 1. The first relation $\sum_1^n (y/\sqrt{n}) = \sqrt{n}.\bar{y}$ projects the point Y on to plane I through the origin in the point $Q(y_1-\bar{y},\ldots,y_n-\bar{y})$, and $OY^2=z_1^2+q_1^2$, where $z_1=\sqrt{n}.\bar{y}$ and $q_1=\sqrt{[\Sigma(y-\bar{y})^2]}$; the second relation $\sum_1^{n-1}(uu'/\sqrt{\Sigma_{n-1}})=\sqrt{\Sigma_{n-1}}.b$ projects Q on to a plane II through the origin (orthogonal to plane I) in the point $Q'(y_1-Y_1,\ldots,y_n-Y_n)$, and $q_1^2=z_2^2+q_2^2$, where $z_2=\sqrt{\Sigma_{n-1}}.b$, and $q_2=\sqrt{[\Sigma(y-Y)^2]}$. Q' is confined to the $(n-2)$ space determined by the intersection of planes I and II. Altogether $OY^2=z_1^2+z_2^2+q_2^2$, and q_2^2 can then be determined as the sum of the squares of $(n-2)$ perpendicular distances from Q' on to $(n-2)$ other hyperplanes through the origin orthogonal to plane II, and mutually orthogonal among themselves. Again an infinity of planes can be chosen for the first, and the one we take is $u_2u_1'=u_1u_2'$, which can if desired be expressed in terms of the original variables x and y as $(x_2-x_3)y_1+(x_3-x_1)y_2+(x_1-x_2)y_3=0$. The mutual independence of the distributions of $\bar{y}\sqrt{n}$, $b\sqrt{[\Sigma(x-\bar{x})^2]}$ and $\Sigma(y-Y)^2$, and their separate distribution functions, then follow.

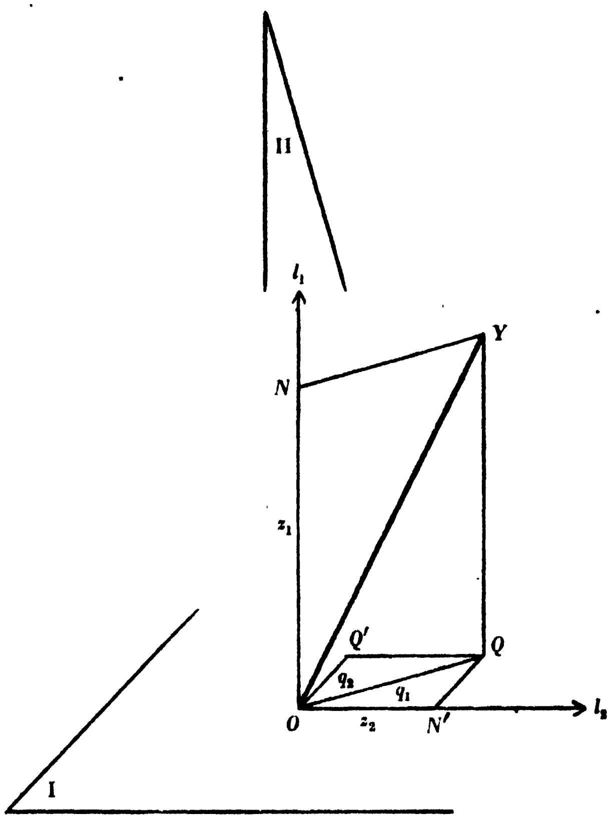


Fig. 1.

REFERENCES

ANDERSON, T. W. (1946). *Ann. Math. Statist.* **17**, 409.
ANDERSON, T. W. & GIRSHICK, M. A. (1944). *Ann. Math. Statist.* **15**, 345.
BARTLETT, M. S. (1933). *Proc. Roy. Soc. Edinb.* **53**, 260.
CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press.
ELFVING, G. (1947). *Skand. AktuarTidskr.* **30**, 56.
FISHER, R. A. (1915). *Biometrika*, **10**, 507.
FISHER, R. A. (1925). *Metron*, **5**, no. 3, p. 90.
FOG, D. (1948). *Biometrika*, **35**, 46.
HSU, P. L. (1939). *Proc. Camb. Phil. Soc.* **35**, 336.
INGHAM, A. E. (1933). *Proc. Camb. Phil. Soc.* **29**, 270.
KENDALL, M. G. (1946). *The Advanced Theory of Statistics*, vol. II. London: Griffin and Co.
MADOW, W. G. (1938). *Trans. Amer. Math. Soc.* **44**, 454.
MAHALANOBIS, P. C., BOSE, R. C. & ROY, S. N. (1937). *Sankhyā*, **3**, 1.
ROMANOVSKY, V. (1925). *Metron*, **5**, no. 4, p. 3.
SIMONSEN, W. (1944, 1945). *Skand. AktuarTidskr.* **27**, 235; **28**, 20.
VAJDA, S. (1945). *Ann. Math. Statist.* **16**, 381.
WILKS, S. S. (1943). *Mathematical Statistics.* Princeton University Press.
WISHART, J. (1928). *Biometrika*, **20A**, 32.
WISHART, J. (1947). *J. Inst. Actu. Stud. Soc.* **7**, 98.
WISHART, J. (1948). *J. Inst. Actu. Stud. Soc.* **8**, 38.
WISHART, J. & BARTLETT, M. S. (1933). *Proc. Camb. Phil. Soc.* **29**, 260.

TESTS OF SIGNIFICANCE IN MULTIVARIATE ANALYSIS

By C. RADHAKRISHNA RAO, *King's College, Cambridge*

CONTENTS

1. Introduction

Attempts have been made in recent years to generalize the univariate analysis of variance technique to the case of multiple variates. The extension of the theory has been slow and only a few methods have been made available for practical use. The starting-point of these researches is the simultaneous sampling distribution of the variances and covariances in samples from a multivariate normal population given by Wishart in 1928. A few years later Hotelling (1931) found the distribution of a quantity T which is a natural extension of Student's distribution to a sample from a multivariate normal population.

Wilks (1932) following the *likelihood ratio method* (Neyman & Pearson, 1928, 1931; Pearson & Neyman, 1930) obtained suitable generalizations in the analysis of variance applicable to several variables. The statistic Λ proposed by him has been found useful in a variety of problems. Bartlett (1934) applied it for testing the significance of treatments with respect to two variables in a varietal trial and indicated its general use in multivariate tests of significance. Wilks (1935) and Hotelling (1935) found it useful in testing the independence of several groups of variates. Recently Plackett (1947) provided an exact test for judging the equality of variances and covariances of various populations, with the use of this criterion. Wilks' statistic supplied some of the basic tests in multivariate analysis but the problem of tabulation has not been tackled except in some limited cases (Wald & Brookner, 1941). A very useful approximation has been suggested by Bartlett (1938) who further demonstrated its use in a paper on 'Multivariate Analysis' read before the Royal Statistical Society in May 1947.

A new line of research was initiated by Fisher (1936) with his introduction of the discriminant function analysis. It has been shown (Barnard, 1935; Martin, 1935; Fairfield Smith, 1936) that a set of multiple measurements may be used to provide a discriminant function *linear in the observations* having the property that, better than *any other linear function*, it will discriminate between any two chosen classes such as taxonomic species, the

two sexes and so on. It has also been shown (Welch, 1939; Rao, 1947) that the linear discriminant function chosen as stated above is the best among all classes of functions not necessarily linear in the set of observations.

The introduction of the discriminant function led to a new method of deriving test criteria suitable for multiple variates. The problem is reduced to the case of a single variate by using a linear compound of the several variables, where the compounding coefficients are chosen so as to maximize the value of a statistic suitable for a single variate. The application of this method to test the differences in mean values for several groups gave rise to the theory of canonical roots of determinantal equations (Roy, 1939; Fisher, 1939; Hsu, 1939*b*). The distribution of the individual roots and the exact nature of tests require further study. Wilks' statistic, which is a symmetric function of the canonical roots, may be considered as providing an overall test of the hypothesis concerned.

The object of the present paper is twofold. The first is to develop a unified approach to the problem of tests of significance in multivariate analysis. The concept of *analysis of dispersion* explained in § 4 (*a*), which is a natural extension of the univariate analysis of variance, has been found useful in discussing multivariate problems. In a recent abstract, Hotelling (1947) developed a method of splitting a generalized measure of dispersion which appears to be different from the method proposed here.

The second is to examine the nature of Bartlett's (1938) approximation and supply appropriate methods for cases needing the exact evaluation of the probabilities.

In presenting the various tests of significance developed in this paper it has been found convenient to consider the problems arising out of a single sample and two samples in the first stage. They depend on simple tests of significance requiring the use of variance ratio tables alone and are of very great importance in practice. The use of Wilks' statistic in multivariate analysis involving more than two samples is considered in the second stage. A number of examples have been worked out to explain the computational procedure.

2. TESTS WITH DISCRIMINANT FUNCTIONS

(*a*) *Two fundamental distributions*

The method of discriminant functions to derive test criteria has been found extremely useful in multivariate analysis. The problem is reduced to the case of a single variable by choosing a linear compound of the variables and constructing a statistic suitable for the case of a single variate. The maximized value of this statistic obtained by a suitable choice of the compounding coefficients is taken as the appropriate test criterion. The distribution of the statistics thus derived in problems involving a single sample and two samples depend on the two fundamental distributions considered below.

Let (w_{ij}), $i, j = 1, 2, \ldots, p$ be the matrix giving the estimates, on n degrees of freedom, of the elements in the dispersion matrix (α_{ij}) of p normally correlated variables. The definition of w_{ij} implies that it has been calculated from a certain sum of products by dividing by the appropriate degrees of freedom. Let $d_1, d_2, \ldots, d_p$ be p normal variates with the same dispersion matrix (α_{ij}) but distributed independently of the w_{ij}'s. Considering only the first r variables $d_1, \ldots, d_r$ the statistic V_r is defined as

$$nV_r = \sum_{i=1}^{r} \sum_{j=1}^{r} w_r^{ij} d_i d_j, \tag{2·1}$$

where (w_r^{ij}) is the matrix reciprocal to (w_{ij}), $i, j = 1, 2, \ldots, r$. An alternative method of calculating this statistic is provided by the equation

$$1 + nV_r = \frac{|w_{ij} + d_i d_j|}{|w_{ij}|} \quad (i, j = 1, 2, \ldots, r).$$

It has been shown by Hotelling (1931) that, when the w_{ij} follow Wishart's distribution (Wishart, 1928) with n degrees of freedom and $E(d_1) = E(d_2) = \ldots = E(d_r) = 0$, the statistic $V_r(n+1-r)/r$ can be used as a variance ratio with r and $(n+1-r)$ degrees of freedom.

The author has shown elsewhere (Rao, 1946*b*) that, if $d_{r+1}, \ldots, d_p$ are distributed independently of $d_1, \ldots, d_r$ and $E(d_{r+1}) = \ldots = E(d_p) = 0$, $E(d_i)$ being not necessarily zero when $i = 1, 2, \ldots, r$, the statistic

$$U = \frac{n+1-p}{p-r}\left\{\frac{1+V_p}{1+V_r} - 1\right\} \tag{2·2}$$

can be used as a variance ratio with $(p-r)$ and $(n+1-p)$ degrees of freedom. The statistic V_p is calculated from the formula (2·1) by using all the p variates.

All the tests of significance considered in this section depend on the use of the statistics defined in (2·1) and (2·2).

(*b*) *Problems of a single sample*

Student's test connected with pairs of observations admits generalization in two directions.

The first is to test whether the means of p correlated variables are the same on the basis of a sample of size N from a p-variate population. When the test shows differences in mean values, there arises the question of deciding whether an assigned contrast involving the p variates differs from the best contrast as determined from the data.

If $x_{1i}, x_{2i}, \ldots, x_{pi}$ are the observations on the ith individual, then they may be replaced by a linear compound $z_i = l_1 x_{1i} + \ldots + l_p x_{pi}$ where the l's satisfy the condition $l_1 + \ldots + l_p = 0$. The problem of determining the best contrast reduces to that of determining the compounding coefficients $l_1, \ldots, l_p$ such that the ratio of mean z to standard deviation of z is a maximum. An alternative method which has some practical advantage is as follows.

By arbitrary choice of constants one can construct $(p-1)$ linear combinations of the variables $x_1, \ldots, x_p$,

$$y_j = m_{1j}x_1 + \ldots + m_{pj}x_p,$$

such that $\sum_i m_{ij} = 0$ for $j = 1, 2, \ldots, (p-1)$. Choosing a linear compound of the x's with coefficients adding to zero is the same as choosing a linear compound of the y's without any restriction on the compounding coefficients. If the linear compound is

$$\lambda_1 y_1 + \lambda_2 y_2 + \ldots + \lambda_{p-1} y_{p-1},$$

then the quantity to be maximized is

$$v = \frac{(\lambda_1 \bar{y}_1 + \ldots + \lambda_{p-1}\bar{y}_{p-1})^2}{\Sigma\Sigma\lambda_i\lambda_j w_{ij}},$$

where

$$w_{ij} = \frac{1}{N-1}\sum_{r=1}^{N}(y_{ir} - \bar{y}_i)(y_{jr} - \bar{y}_j).$$

Observing that only the ratios of λ's are uniquely determinable the equations giving λ's may be written as $\lambda_1 w_{1i} + \ldots + \lambda_{p-1} w_{p-1i} = \bar{y}_i \quad (i = 1, 2, \ldots, (p-1))$,

with the solution as $\lambda_i = w^{1i}\bar{y}_1 + \ldots + w^{p-1i}\bar{y}_{p-1} \quad (i = 1, 2, \ldots, (p-1))$,

where the matrix (w^{ij}) is reciprocal to (w_{ij}). This supplies the best linear compound of the y's which on transformation to the x's gives the best contrast determinable from the data.

The maximum value of v is given by

$$\Sigma\lambda_i\bar{y}_i = \Sigma\Sigma w^{ij}\bar{y}_i\bar{y}_j.$$

If $V_{p-1} = N(\Sigma\Sigma w^{ij}\bar{y}_i\bar{y}_j)/(N-1)$, then, on the hypothesis that all the x's have the same mean value, the conditions required for the use of the statistic (2·1) are satisfied so that

$$V_{p-1}(N-p+1)/(p-1)$$

can be used as the variance ratio with $(p-1)$ and $(N-p+1)$ degrees of freedom to test the above hypothesis.

The statistic V_{p-1} is invariant for all sets of coefficients chosen to construct the y's from the x's so that in any practical problem either conveniently or conventionally chosen linear contrasts of the x's may be used to define the y's.

The test used above is essentially the one given by Hotelling (1931). The point of interest is to show how the test is derivable by the method of discriminant functions involving some restrictions on the compounding coefficients. Also the author is not aware whether the use of Hotelling's T in testing the equality of means of p correlated variables has been explicitly mentioned anywhere.

To test whether the best contrast as determined from the data is in agreement with an assigned contrast $\xi_1 x_1 + \ldots + \xi_p x_p$ or $\eta_1 y_1 + \ldots + \eta_{p-1} y_{p-1}$ in terms of the y's one may proceed as follows.

The appropriate statistic for testing the significance of the assigned contrast is

$$V_1 = \frac{N(\eta_1\bar{y}_1 + \ldots + \eta_{p-1}\bar{y}_{p-1})^2}{(N-1)(\Sigma\Sigma\eta_i\eta_j w_{ij})},$$

where $V_1(N-1)$ is the variance ratio with 1 and $(N-1)$ degrees of freedom. The appropriate statistic for all the $(p-1)$ contrasts is V_{p-1} as considered before. The hypothesis specifies that all contrasts orthogonal to the assigned one have zero mean so that the conditions for the use of the statistic (2·2) are satisfied. Hence

$$U = \frac{(N-p+1)}{(p-2)}\left\{\frac{1+V_{p-1}}{1+V_1} - 1\right.$$

can be used as the variance ratio with $(p-2)$ and $(N-p+1)$ degrees of freedom to test the above hypothesis.

The above test can be generalized to answer the problem whether a set of k assigned contrasts contain the best contrast. In this case the statistic

$$U = \frac{(N-p+1)}{(p-k-1)}\left\{\frac{1+V_{p-1}}{1+V_k} - 1\right\}$$

can be used as the variance ratio with $(p-k-1)$ and $(N-p+1)$ degrees of freedom.

The second generalization of Student's t is concerned with testing, on the basis of a sample of size N from a $2p$-variate population containing the variables $y_1, \ldots, y_{2p}$ whether $E(y_i) = E(y_{i+p})$ for $i = 1, 2, \ldots, p$.

From the $2p$ variates $y_1, \ldots, y_{2p}$ one can construct the p variates $z_i = y_i - y_{i+p}$, $i = 1, 2, \ldots, p$ in which case the problem reduces to that of testing the hypothesis $E(z_i) = 0$, $i = 1, 2, \ldots, p$. The variance ratio with p and $(N-p)$ degrees of freedom to test the above hypothesis is

$$\frac{N(N-p)}{p(N-1)}\Sigma\Sigma w^{ij}\bar{z}_i\bar{z}_j,$$

where (w^{ij}) is reciprocal to (w_{ij}) giving the estimates of the variances and covariances of the z's.

This test may be useful in biometry where the asymmetry of organisms is considered. The sets of variables $y_1, ..., y_p$ and $y_{p+1}, ..., y_{2p}$ will then correspond to measurements on the right and left sides of an organism.

Example 1. The data of Table 1 consist of weights of cork borings taken by the author from the north (N), east (E), south (S) and west (W) directions of the trunk for 28 trees in a block of plantations. The problem is to test whether the bark deposit varies in thickness and hence in weight in the four directions. It was suggested by Prof. Mahalanobis that the bark deposit is likely to be uniform in N and S directions and also uniform but less in E and W directions, so that N − E − W + S can be taken as the best contrast. This can, however, be tested from the given data as shown below.

Table 1. *Weights of cork borings (in centigrams) in the four directions for* 28 *trees*

N	E	S	W	N	E	S	W
72	66	76	77	91	79	100	75
60	53	66	63	56	68	47	50
56	57	64	58	79	65	70	61
41	29	36	38	81	80	68	58
32	32	35	36	78	55	67	60
30	35	34	26	46	38	37	38
39	39	31	27	39	35	34	37
42	43	31	25	32	30	30	32
37	40	31	25	60	50	67	54
33	29	27	36	35	37	48	39
32	30	34	28	39	36	39	31
63	45	74	63	50	34	37	40
54	46	60	52	43	37	39	50
47	51	52	43	48	54	57	43

It has been found in similar studies that there exists a significant correlation between contrasts such as (N − E) and (S − W) so that the method of fitting constants for the four directions and the individual trees by the method of least squares is not appropriate. The three contrasts arising out of the four weights may, however, be treated as three correlated variables in which case the theory developed above is applicable.

It is interesting to observe that the individual weights in Table 1 are exceedingly asymmetrically distributed.* This does not, however, invalidate the test so long as the contrasts are normally distributed. In fact, the distribution of the individual weights depends on the nature of the plants and the variation between plants. If the above condition is satisfied, it is not necessary that the individual weights should follow any distribution law of the known type. It may be sometimes necessary to make a transformation (such as log, square or cube root) of the variables under consideration to ensure that the contrasts of the transformed variables are symmetrically distributed if the contrasts of the original variables are not so.

As observed earlier the contrasts may be conveniently or conventionally chosen. For the above example one may choose the simple set of contrasts

$$y_1 = \mathrm{N} - \mathrm{E} - \mathrm{W} + \mathrm{S}, \quad y_2 = \mathrm{S} - \mathrm{W}, \quad y_3 = \mathrm{N} - \mathrm{S}.$$

* I am indebted to Prof. E. S. Pearson for drawing my attention to this fact.

The mean values and estimates of variances and covariances based on 27 degrees of freedom for the y's are $\bar{y}_1 = 8{\cdot}8571,\quad \bar{y}_2 = 4{\cdot}5000,\quad \bar{y}_3 = 0{\cdot}8571,$

$$(w_{ij}) = \begin{pmatrix} 128{\cdot}7200 & 61{\cdot}4076 & -21{\cdot}0211 \\ 61{\cdot}4076 & 56{\cdot}9259 & -28{\cdot}2963 \\ -21{\cdot}0211 & -28{\cdot}2963 & 63{\cdot}5344 \end{pmatrix}.$$

The coefficients in the best linear function $\lambda_1 y_1 + \lambda_2 y_2 + \lambda_3 y_3$ are given by the equations

$$128{\cdot}7200\lambda_1 + 61{\cdot}4076\lambda_2 - 21{\cdot}0211\lambda_3 = 8{\cdot}8571,$$
$$61{\cdot}4076\lambda_1 + 56{\cdot}9259\lambda_2 - 28{\cdot}2963\lambda_3 = 4{\cdot}5000,$$
$$-21{\cdot}0211\lambda_1 - 28{\cdot}2963\lambda_2 + 63{\cdot}5344\lambda_3 = 0{\cdot}8571.$$

Solving, one gets $\lambda_1 = 0{\cdot}05620,\quad \lambda_2 = 0{\cdot}04415,\quad \lambda_3 = 0{\cdot}05174,$

so that the best contrast is

$$\lambda_1(\mathrm{N-E-W+S}) + \lambda_2(\mathrm{S-W}) + \lambda_3(\mathrm{N-S})$$
$$= 0{\cdot}10794\mathrm{N} - 0{\cdot}05620\mathrm{E} - 0{\cdot}10035\mathrm{W} + 0{\cdot}04861\mathrm{S},$$

or

$$1{\cdot}0794\mathrm{N} - 0{\cdot}5620\mathrm{E} - 1{\cdot}0035\mathrm{W} + 0{\cdot}4861\mathrm{S},$$

obtained by multiplying the coefficients by 10 (arbitrarily).

The statistic for testing the hypothesis of equality of means is

$$V_{p-1} = \frac{N}{N-1}(\lambda_1\bar{y}_1 + \lambda_2\bar{y}_2 + \lambda_3\bar{y}_3)$$
$$= \tfrac{28}{27}(0{\cdot}740790) = 0{\cdot}768226.$$
$$V_{p-1}(N-p+1)/(p-1) = 0{\cdot}768226(28-4+1)/3 = 6{\cdot}4019.$$

The quantity $6{\cdot}4019$ as the variance ratio with 3 and 25 degrees of freedom is significant at 1% level so that the bark deposit cannot be considered uniform in the four directions.

The assigned contrast is represented by y_1 and to test for its significance one can construct the statistic

$$V_1 = \frac{N}{N-1}\frac{\bar{y}_1^2}{w_{11}} = \frac{28}{27}\frac{(8{\cdot}8571)^2}{128{\cdot}7200} = 0{\cdot}632020.$$

The quantity $(N-1)V_1 = 17{\cdot}0645$ as the variance ratio with 1 and 27 degrees of freedom is highly significant.

To test whether the assigned contrast agrees with that estimated from the data, the statistic U defined in (2·2) has to be calculated.

$$U = \frac{N-p+1}{p-2}\left\{\frac{1+V_{p-1}}{1+V_1} - 1\right\} = \frac{25}{2}\left\{\frac{1{\cdot}768226}{1{\cdot}632020} - 1\right\} = 1{\cdot}0431.$$

This value of U as the variance ratio with 2 and 25 degrees of freedom is small so that the evidence supplied by the data is not sufficient to reject the assigned contrast as not the best, although the ratios of the coefficients in the estimated contrast depart considerably from those assigned.

(c) *Mahalanobis' D^2 and problems of two samples*

Let N_1 and N_2 be the sample sizes from two populations π_1 and π_2 characterized by $(p+q)$ variates. The sample means for the ith character are represented by $\bar{x}_{i1}$ and $\bar{x}_{i2}$ for π_1 and π_2 respectively. The estimated value of the covariance is given by

$$(N_1+N_2-2)\,w_{ij} = \sum_{t=1}^{N_1}(x_{i1t}-\bar{x}_{i1})(x_{j1t}-\bar{x}_{j1}) + \sum_{t=1}^{N_2}(x_{i2t}-\bar{x}_{i2})(x_{j2t}-\bar{x}_{j2}).$$

Mahalanobis' (1936) distance between the two populations as estimated from the sample on the basis of the first p characters is

$$D_p^2 = \sum_{i=1}^{p} \sum_{j=1}^{p} w_p^{ij} (\bar{x}_{i1} - \bar{x}_{i2})(\bar{x}_{j1} - \bar{x}_{j2}),$$

where (w_p^{ij}) is reciprocal to (w_{ij}), $i, j = 1, 2, \ldots, p$. The exact distribution of D_p^2 on the hypothesis specifying real differences in mean values has been given by Bose & Roy (1938). To test the hypothesis specifying no difference in mean values of the p characters for π_1 and π_2 the statistic

$$\frac{N_1 N_2 (N_1 + N_2 - p - 1)}{p(N_1 + N_2)(N_1 + N_2 - 2)} D_p^2$$

can be used as the variance ratio with p and $(N_1 + N_2 - 1 - p)$ degrees of freedom. The appropriate distance function based on $(p+q)$ characters is

$$D_{p+q}^2 = \Sigma\Sigma w_{p+q}^{ij} (\bar{x}_{i1} - \bar{x}_{i2})(\bar{x}_{j1} - \bar{x}_{j2}).$$

To test whether the q additional characters lead to significant differences between π_1 and π_2, independently of the first p characters, a comparison may be made of the magnitudes of D_p^2 and D_{p+q}^2. One may choose the ratio

$$R = \frac{1 + N_1 N_2 D_{p+q}^2 / (N_1 + N_2)(N_1 + N_2 - 2)}{1 + N_1 N_2 D_p^2 / (N_1 + N_2)(N_1 + N_2 - 2)}$$

which has some theoretical advantage as observed in the next section. The conditions for the use of the statistic defined in (2·2) are satisfied so that

$$U = \frac{N_1 + N_2 - p - q - 1}{q} (R - 1)$$

is the variance ratio with q and $(N_1 + N_2 - p - q - 1)$ degrees of freedom.

It has been shown by Fisher (1938) that the test with Mahalanobis' D^2 is equivalent to testing the difference in the mean values of the discriminant function as estimated from the samples. Apart from tests of significance, the discriminant function is used for the purpose of assigning an individual to its proper class. This leads us to the problem of determining the number and nature of characters used in order that the errors of classification may not be large. It is, however, to be expected that the errors of classification will decrease as the number of characters is increased. On the other hand, since the various characters are correlated with one another the addition of some characters to a basic panel may not reduce or at any rate substantially decrease the errors. In this case, it may not be worth while to increase the number of characters so that the numerical computation involved may not be heavy. A test designed to judge the significance of the reduction in errors of classification by the addition of some new characters is valuable in problems of this nature. Such a test is mathematically equivalent to a test of differences in mean values of the new variables after eliminating the differences in the basic characters. If p is the number of basic characters and q the number of additional characters the test may also be conceived as a comparison of Mahalanobis' distances based on p and $(p+q)$ characters.

If it is desired to test whether an assigned discriminant function $\eta_1 x_1 + \ldots + \eta_p x_p$ of p characters differs from that calculated from the data, then the statistic

$$\frac{N_1 + N_2 - p - 1}{p - 1} \left\{ \frac{1 + N_1 N_2 D_p^2 / (N_1 + N_2)(N_1 + N_2 - 2)}{1 + N_1 N_2 D_1^2 / (N_1 + N_2)(N_1 + N_2 - 2)} - 1 \right\}$$

where $D_1^2 = \Sigma\Sigma \eta_i \eta_j (\bar{x}_{i1} - \bar{x}_{i2})(\bar{x}_{j1} - \bar{x}_{j2}) / \Sigma\Sigma \eta_i \eta_j w_{ij}$, can be used as the variance ratio with $(p-1)$ and $(N_1 + N_2 - p - 1)$ degrees of freedom (Fisher, 1940; Bartlett, 1939).

Example 2. The following Tables 2 and 3 reproduced from Fisher (1938) give the mean values based on fifty observations each and the covariances based on $50+50-2$ degrees of freedom for four characters in two species of plants *Iris versicolor* and *Iris setosa*.

The solution of the equations

$$\lambda_1 w_{1i} + \ldots + \lambda_4 w_{4i} = d_i \quad (i = 1, 2, 3, 4)$$

is obtained as

$$\lambda_1 = -3{\cdot}0528, \quad \lambda_2 = -18{\cdot}0229, \quad \lambda_3 = 21{\cdot}7662, \quad \lambda_4 = 30{\cdot}8442,$$

so that the discriminant function is

$$-3{\cdot}0528x_1 - 18{\cdot}0229x_2 + 21{\cdot}7662x_3 + 30{\cdot}8442x_4.$$

The value of D_4^2 is* $\qquad \lambda_1 d_1 + \ldots + \lambda_4 d_4 = 103{\cdot}2335.$

To test for the differences in mean values one can use the statistic

$$\frac{N_1 N_2(N_1+N_2-1-4)}{(N_1+N_2)(N_1+N_2-2)} \frac{D_4^2}{4} = \frac{50\times 50\times 95}{100\times 98\times 4}\times 103{\cdot}2335$$
$$= \tfrac{95}{4}(26{\cdot}3350) = 625{\cdot}4515,$$

which as the variance ratio with 4 and 95 degrees of freedom is highly significant.

Table 2. *Observed mean values based on fifty observations each for the two species*

Character	*Iris versicolor*	*Iris setosa*	Difference (d)
Sepal length (x_1)	5·936	5·006	0·930
Sepal width (x_2)	2·770	3·428	−0·658
Petal length (x_3)	4·260	1·462	2·798
Petal width (x_4)	1·326	0·246	1·080

Table 3. *The pooled covariance matrix based on* 98 *degrees of freedom*

	x_1	x_2	x_3	x_4
x_1	0·195340	0·092200	0·099626	0·033055
x_2	0·092200	0·121079	0·047175	0·025251
x_3	0·099626	0·047175	0·125488	0·039586
x_4	0·033055	0·025251	0·039586	0·025106

If only the lengths are considered (i.e. x_1 and x_3) the equations leading to the discriminant function are

$$0{\cdot}195340\mu_1 + 0{\cdot}099626\mu_2 = 0{\cdot}930,$$
$$0{\cdot}099626\mu_1 + 0{\cdot}125488\mu_2 = 2{\cdot}798,$$

so that $\mu_1 = -11{\cdot}1088$, $\mu_2 = 31{\cdot}1163$ and

$$D_2^2 = \mu_1(0{\cdot}930) + \mu_2(2{\cdot}798) = 76{\cdot}7322.$$

To answer the question whether the widths (x_2 and x_4) supply independent information the significance of the ratio

$$R = \frac{1 + N_1 N_2 D_4^2/(N_1+N_2)(N_1+N_2-2)}{1 + N_1 N_2 D_2^2/(N_1+N_2)(N_1+N_2-2)}$$
$$= \frac{1+26{\cdot}3350}{1+19{\cdot}5745} = 1{\cdot}3286$$

* This method of calculating D^2 avoids the actual inversion of the matrix (w_{ij}). The λ coefficients obtained by this process can be directly used in the construction of the discriminant function.

has to be tested. The value of the statistic

$$U = (R-1)(N_1+N_2-1-4)/2 = 95(0{\cdot}3286)/2 = 15{\cdot}6085$$

as the variance ratio with 2 and 95 degrees of freedom is significant at 1 % level. This shows that the widths in association with the lengths lead to further discrimination of the species, so that there is a significant reduction in the errors of classification.

(d) *Test for the equality of discriminant functions*

If four samples of sizes N_1, N_2, N_3 and N_4 from populations π_1, π_2, π_3 and π_4 are available one can test whether the discriminant functions between π_1, π_2 and π_3, π_4 are significantly different by an extension of the test criterion discussed above. It is a necessary condition of the test that the variances and covariances are identical in the four populations π_1, π_2, π_3 and π_4. No reasonably simple test can be constructed to establish the equivalence of the discriminant functions when this condition is not satisfied.

Let (w_{ij}) be the dispersion matrix based on $(N_1+N_2+N_3+N_4-4)$ degrees of freedom. If $d_1, \ldots, d_p$ are the differences in mean values for π_1 and π_2 and $d'_1, \ldots, d'_p$ are those for π_3 and π_4, the test for equality of discriminant functions is identical with the testing of the hypothesis

$$E(d_i) = E(d'_i) \quad (i = 1, 2, \ldots, p),$$

or

$$E(d_i) = E(-d'_i) \quad (i = 1, 2, \ldots, p).$$

The variance ratios with p and $n = (N_1+N_2+N_3+N_4-3-p)$ degrees of freedom for the two cases are

$$\frac{n}{p}\frac{f(N)}{n+p-1}\Sigma\Sigma w^{ij}(d_i-d'_i)(d_j-d'_j)$$

and

$$\frac{n}{p}\frac{f(N)}{n+p-1}\Sigma\Sigma w^{ij}(d_i+d'_i)(d_j+d'_j)$$

where

$$\frac{1}{f(N)} = \left(\frac{1}{N_1}+\frac{1}{N_2}+\frac{1}{N_3}+\frac{1}{N_4}\right).$$

The equality of discriminant functions is indicated if at least one of the statistics is not significant. Similar tests can be constructed for judging the differences in discriminant functions in parallel samples from two populations or between π_1, π_2 and π_1, π_3.

3. Generalization of D^2 and the large sample theory for several groups

Let there be k multivariate populations $\pi_1, \pi_2, \ldots, \pi_k$ from which samples of sizes $N_1, N_2, \ldots, N_k$ are available for $p+q$ characters. The common covariance matrix assumed to be known or estimated on a large number of degrees of freedom is represented by $(\alpha_{ij}^{(p)})$ for the first p characters and by $(\alpha_{ij}^{(p+q)})$ for all the $(p+q)$ characters. The inverse of $(\alpha_{ij}^{(p)})$ is represented by $(\alpha^{ij}_{(p)})$ and that of $(\alpha_{ij}^{(p+q)})$ by $(\alpha^{ij}_{(p+q)})$. Let $\bar{x}_{i1}, \bar{x}_{i2}, \ldots$ be the mean values of the ith character in the first, second, etc., populations.

It has been shown by Hsu (1939*a*) and Rao (1945) that the statistic,

$$V_{p,k} = \sum_{i,j=1}^{p}\sum \alpha^{ij}_{(p)} \sum_{r=1}^{k} N_r(\bar{x}_{ir}-\bar{x}_{i.})(\bar{x}_{jr}-\bar{x}_{j.}),$$

where $\bar{x}_{i.} = (\Sigma N_r \bar{x}_{ir})/(\Sigma N_r)$, can be used as χ^2 with $p(k-1)$ degrees of freedom to test the hypothesis that the mean values are the same in all the k populations for these p characters. The statistic V_{pk} is a suitable generalization of the Mahalanobis' D^2 in its classical form and its theoretical derivation has been discussed by the author (Rao, 1945).

When this test indicates differences in mean values it is, in some problems, necessary to test whether the observations on q additional characters supply independent information for discrimination. The statistic for testing the differences in means for all the $p+q$ characters is

$$V_{p+q,k} = \sum_{i,j=1}^{p+q}\sum \alpha_{(p+q)}^{ij} \sum_{r=1}^{k} N_r(\bar{x}_{ir}-\bar{x}_{i.})(\bar{x}_{jr}-\bar{x}_{j.}),$$

which can be used as a χ^2 with $(p+q)(k-1)$ degrees of freedom. The $q(k-1)$ additional degrees of freedom bring in the contribution

$$V_{p+q,k}-V_{p,k},$$

and the significance of this difference can be appropriately used to judge the significance of the information supplied by the additional characters. This difference can be used as a χ^2 with $q(k-1)$ degrees of freedom as shown below.

The hypothesis that the new characters do not lead to further discrimination of the populations specifies that any linear function of the $(p+q)$ characters uncorrelated with each of the p characters has the same mean value for all the k populations. There are q such linear functions and treating them as q variables a χ^2 with $q(k-1)$ degrees of freedom can be constructed to test the above hypothesis. The above method of taking the difference is only an alternative way of calculating this χ^2. For $V_{p+q,k}$ calculated from all the $(p+q)$ characters, being invariant under linear transformations of the variables, is equal to $V_{p,k}+\chi^2$ calculated from the p original characters and the q linear functions chosen to be uncorrelated with each of the p characters.

In the above derivation it has been assumed that the variances and covariances are known and the distributions are asymptotically correct when they are estimated on a large number of degrees of freedom. When more than two populations are involved the pooled estimates of the covariances have, usually, a sufficiently large number of degrees of freedom to validate the use of the asymptotic distributions. More exact tests for cases involving small numbers of degrees of freedom are given in the next section.

4. TESTS WITH WILKS' Λ CRITERION

(a) *Analysis of dispersion and the theoretical aspects of the Λ criterion*

In the univariate analysis of variance tests of significance reduce to the comparison of two independently distributed mean squares. One of the mean squares is an unbiased estimate of the variance to which a single observation in any particular class is subject and is called the error variance. The other is only so when the null hypothesis which is being tested is correct and may be called the mean square due to deviation from the hypothesis. The test depends only on the individual degrees of freedom of two mean squares.

When each observation consists of p mutually correlated variables there are p total sums of squares and $p(p-1)/2$ total sums of products which can be analysed into various categories. This process which involves the technique of analysing the variances and covariances of multiple correlated variables may be termed the *analysis of dispersion.** The term

* Prof. R. A. Fisher suggested that this method can be described as either analysis of covariance or analysis of dispersion. Since the term covariance analysis is conventionally used in problems of adjustment for concomitant variation, I have used the term analysis of dispersion to cover a wider variety of problems considered in this article.

dispersion has been used by Prof. Mahalanobis to indicate the scatter of a set of observations as measured by the variances and covariances. Following this terminology the total dispersion may be said to be analysed into dispersion due to various categories.

If we represent the total sum of products by the matrix $S = (S_{ij})$, then the analysis of dispersion consists in analysing each element such as S_{ij}, according to the usual procedure, into various categories with the corresponding distribution of degrees of freedom. The dispersion due to any category supplies the sum of products (denoted for brevity by S.P.) matrix which on division by the degrees of freedom gives the mean product (denoted by M.P.) matrix. The S.P. matrix leading to unbiased estimates of the variances and covariances to which a single set of variables is subject is called the S.P. matrix due to error. This error matrix may be denoted by W with w as its degrees of freedom. In the analysis of dispersion the S.P. matrix due to any other category leads to unbiased estimates of variances and covariances only when the null hypothesis regarding that category is true. This may be called the S.P. matrix due to deviation from the hypothesis. If such a matrix is represented by Q with q as its degrees of freedom, then the problem of testing the null hypothesis consists in comparing the matrices $w^{-1}W$ and $q^{-1}Q$. The simultaneous comparison of the estimates of the variances and covariances appears to be a natural extension of the comparison of variances in the case of a single variate.

The appropriate test criteria for comparison may be obtained by extending the method of discriminant function analysis. A linear compound of the variables is taken and the compounding coefficients are chosen such that the ratio of mean squares due to deviation from hypothesis and due to error for this variable is a maximum. The ratio F^2 which comes out as a root of the determinantal equation

$$Q - \frac{q}{w} F^2 W = 0$$

may be used as the appropriate test criterion. If $| W | \neq 0$, the number of non-zero roots of this equation is equal to the number of variables under consideration or q, the number of degrees of freedom of Q, whichever is smaller. An adequate comparison of $w^{-1}W$ and $q^{-1}Q$ must involve the tests of significance of all the roots. If $F_1^2, F_2^2, \ldots$ represent the various roots, it is easy to verify that

$$\left(1 + \frac{q}{w} F_1^2\right)\left(1 + \frac{q}{w} F_2^2\right) \ldots = \frac{| W + Q |}{| W |}.$$

The ratio $| W |/| W + Q |$ denoted by Λ decreases as the roots increase and a significantly small value of Λ may be taken as providing the significance of one or more of the roots. This is the underlying theory of the Λ criterion arrived at by Wilks (1932) by using the likelihood ratio method and later extended by Bartlett (1934)* for general use in multivariate analysis.

This, however, does not provide a satisfactory test, for when only one or a smaller number of roots than the total indicate real differences, their significance may be obscured by the use of the overall test. Its use can be recommended only in situations where small deviations from the hypothesis can be ignored.

* In a paper read before the Royal Statistical Society in May 1947 (Bartlett 1947), Dr Bartlett suggested a method of factorizing Λ, arising out of a category in the analysis of dispersion, which appeared different from the procedure I have outlined above. In my discussion of Bartlett's paper, I pointed out the difference between his approach in some problems and the general approach by the method of analysis of dispersion, which alone I think can lead to unbiased tests of significance. Such a factorization leads to a valid test in the case $q = 1$ as I have shown elsewhere (Rao, 1946*b*, p. 409, equation 2·14). But this is not true in general.

(b) *The distribution of Λ and its practical use*

The following notation will be used throughout this and the subsequent sections.

Analysis of dispersion for p variables

Due to	D.F.	S.P. matrix
(1) Deviation from hypothesis	q	Q
(2) Error	$n-q$	W
(3) Total	n	$W+Q$

$$\Lambda = |W|/|W+Q|$$

If the number of variables involved is p, then assuming that the elements of W are distributed independently of those of Q it is easy to derive the tth moment of Λ (Wilks, 1932; Bartlett, 1934) as

$$\prod_{i=0}^{p-1} \frac{\Gamma\{\frac{1}{2}(n-i)\}\,\Gamma\{\frac{1}{2}(n-q-i)+t\}}{\Gamma\{\frac{1}{2}(n-q-i)\}\,\Gamma\{\frac{1}{2}(n-i)+t\}}.$$

The tests based on the exact distributions given by Wilks (1932) and Nair (1939) for some particular cases are reproduced below.

	Nature of test	
	Variance ratio	Degrees of freedom
$q=1$, for any p	$\dfrac{1-\Lambda}{\Lambda}\dfrac{n-p}{p}$	p and $(n-p)$
$q=2$, for any p	$\dfrac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\dfrac{n-p-1}{p}$	$2p$ and $2(n-p-1)$
$p=1$, for any q	$\dfrac{1-\Lambda}{\Lambda}\dfrac{n-q}{q}$	q and $(n-q)$
$p=2$, for any q	$\dfrac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\dfrac{n-q-1}{q}$	$2q$ and $2(n-q-1)$

For other values of p and q, the exact values of the probabilities can be obtained by the use of the χ^2 tables alone as shown below.

It has been shown by Wald & Brookner (1941) that the distribution of the statistic $v = -\log\Lambda$ is of the form

$$2^{\frac{1}{2}pq}C(n)\,e^{-\frac{1}{2}nv}\,v^{\frac{1}{2}pq-1}\left(\sum_{s=0}^{\infty}\frac{\beta_s v^s}{2^{s+\frac{1}{2}pq}\,\Gamma(s+\frac{1}{2}pq)}\right)dv,$$

where

$$C(n) = \prod_{i=0}^{p-1}\frac{\Gamma\{\frac{1}{2}(n-i)\}}{\Gamma\{\frac{1}{2}(n-q-i)\}},$$

and the β's are the coefficients in the expansion of $\dfrac{(n/2)^{\frac{1}{2}pq}}{C(n)}$ in powers of $1/n$,

$$\frac{(n/2)^{\frac{1}{2}pq}}{C(n)} = \beta_0+\frac{\beta_1}{n}+\frac{\beta_2}{n^2}+\dots.$$

For the purposes of examining Bartlett's approximation and obtaining a quickly convergent series for the tabulation of percentage points, the transformation

$$V = \left(n - \frac{p+q+1}{2}\right) v = mv$$

is made. Changing over to V from v the above distribution becomes

$$2^{\frac{1}{2}pq} D(m)\, m^{-\frac{1}{2}pq} e^{-\frac{1}{2}V} V^{\frac{1}{2}pq-1} \left(\sum_{s=0}^{\infty} \frac{\gamma_s V^s}{m^s 2^{s+\frac{1}{2}pq} \Gamma(s+\frac{1}{2}pq)} \right) dV,$$

where
$$D(m) = \prod_{i=0}^{p-1} \frac{\Gamma\left\{\frac{1}{2}\left(m + \frac{p+q+1}{2} - i\right)\right\}}{\Gamma\left\{\frac{1}{2}\left(m + \frac{p-q+1}{2} - i\right)\right\}}.$$

It is easy to recognize that the γ's depend only on p and q and that they are the coefficients in the expansion of $\dfrac{(m/2)^{\frac{1}{2}pq}}{D(m)}$ in powers of $1/m$.

The asymptotic expansion of $\dfrac{(m/2)^{\frac{1}{2}pq}}{D(m)}$ can be calculated from the formulae

$$\left(\frac{m+2}{m}\right)^{\frac{1}{2}pq} \frac{D(m)}{D(m+2)} = \frac{\gamma_0 + \frac{\gamma_1}{m+2} + \dots}{\gamma_0 + \frac{\gamma_1}{m} + \dots},$$

and
$$\frac{D(m)}{D(m+2)} = \prod_{i=0}^{p-1} \frac{m - i + (p-q+1)/2}{m - i + (p+q+1)/2}.$$

Taking the logarithms of the first equality one gets

$$\log\left(\gamma_0 + \frac{\gamma_1}{m+2} + \dots\right) = \frac{pq}{2} \log\left(1 + \frac{2}{m}\right) + \log\left(\gamma_0 + \frac{\gamma_1}{m} + \dots\right)$$
$$+ \sum_i \log\left(1 + \frac{p-q+1-2i}{2m}\right) - \sum_i \log\left(1 + \frac{p+q+1-2i}{2m}\right).$$

Expanding in powers of $(1/m)$ and equating coefficients of like powers on both sides, the first five γ's come out as

$$\gamma_0 = 1,^*$$
$$\gamma_1 = \frac{pq}{2} + \frac{1}{16} \sum_{i=0}^{p-1} \{(p-q+1-2i)^2 - (p+q+1-2i)^2\}$$
$$= 0,$$
$$\gamma_2 = -\frac{pq}{3} + \frac{1}{96} \sum_{i=0}^{p-1} \{(p+q+1-2i)^3 - (p-q+1-2i)^3\}$$
$$= -\frac{pq}{3} + \frac{pq}{48}(p^2+q^2+11)$$
$$= \frac{pq}{48}(p^2+q^2-5),$$

* That the value of $\gamma_0 = 1$ is easily seen by making $m \to \infty$ in the distribution of V in which case it is reduced to the χ^2 form. The comparison of the coefficients, in fact, give the values of the ratios γ_1/γ_0, γ_2/γ_0, ...

$$\gamma_3 = 2\gamma_2 + \frac{pq}{3} + \frac{1}{384}\sum_{i=0}^{p-1}\{(p-q+1-2i)^4-(p+q+1-2i)^4\}$$

$$= 2\gamma_2 + \frac{pq}{3} - \tfrac{16}{384}pq(p^2+q^2+3)$$

$$= 0,$$

$$\gamma_4 = -4\gamma_2 + \frac{\gamma_2^2}{2} - \tfrac{2}{5}pq + \frac{1}{1280}\sum_{i=0}^{p-1}\{(p+q+1-2i)^5-(p-q+1-2i)^5\}$$

$$= -4\gamma_2 + \frac{\gamma_2^2}{2} - \tfrac{2}{5}pq + \frac{pq}{1920}\{3p^4+3q^4+110(p^2+q^2)+10p^2q^2+127\}$$

$$= \frac{\gamma_2^2}{2} + \frac{pq}{1920}\{3p^4+3q^4+10p^2q^2-50(p^2+q^2)+150\}.$$

Considering only terms up to the fourth power in $(1/m)$ the distribution of V becomes

$$\left(\frac{2}{m}\right)^{\frac{1}{2}pq} D(m)\, e^{-\frac{1}{2}V} V^{\frac{1}{2}pq-1}\left\{\frac{1}{2^{\frac{1}{2}pq}\Gamma(\frac{1}{2}pq)} + \frac{\gamma_2}{m^2}\frac{V^2}{2^{2+\frac{1}{2}pq}\Gamma(2+\frac{1}{2}pq)} + \frac{\gamma_4}{m^4}\frac{V^4}{2^{4+\frac{1}{2}pq}\Gamma(4+\frac{1}{2}pq)}\right\}.$$

The probability of V exceeding an observed value V_0, then, becomes

$$\left(\frac{2}{m}\right)^{\frac{1}{2}pq} D(m)\left\{P_{pq} + \frac{\gamma_2}{m^2}P_{pq+4} + \frac{\gamma_4}{m^4}P_{pq+8} + \dots\right\}$$

where

P_{pq+s} = Probability of χ^2 with $pq+s$ degrees of freedom exceeding the value V_0.

One may go a step further and expand $\left(\frac{2}{m}\right)^{\frac{1}{2}pq} D(m)$ in powers of $(1/m)$ in which case the above series becomes

$$P_{pq} + \frac{\gamma_2}{m^2}(P_{pq+4}-P_{pq}) + \frac{1}{m^4}\{\gamma_4(P_{pq+8}-P_{pq}) - \gamma_2^2(P_{pq+4}-P_{pq})\} + \dots.$$

This form is most convenient for the calculation of the required probability. The quantities γ_2 and γ_4 are simple functions of p and q only. Using χ^2 tables, the γ's and powers of $(1/m)$, the probability of V exceeding V_0 can be calculated to a sufficient degree of accuracy.

Bartlett (1938) suggested the use of

$$V = -m\log_e \Lambda = -\left(n - \frac{p+q+1}{2}\right)\log_e \Lambda$$

as χ^2 with pq degrees of freedom. This corresponds to using the first term of the series. Since the second term is $o(1/m^2)$, its contribution is very small even for moderately large m so that, in many practical problems, Bartlett's approximation can be safely used. For small values of m one may use the second and the third terms depending on the accuracy needed in any problem.

(c) *Test of differences in mean values for several populations*

Let $\pi_1, \dots, \pi_k$ be k populations from which samples of sizes $N_1, \dots, N_k$ for p correlated variables are available. The dispersion has to be analysed into 'between' and 'within' populations. The S.P. matrix 'within' populations (the error) has $N_1 + \dots + N_k - k$ degrees of freedom and that 'between' populations has $(k-1)$ degrees of freedom. If these are

represented by W and Q then the statistic to be used for testing the differences in mean values is

$$V = -m \log_e \Lambda,$$

where

$$\Lambda = |W|/|W+Q|,$$
$$m = n - \frac{p+q+1}{2},$$
$$n = N_1 + \ldots + N_k - 1,$$
$$q = k-1.$$

The exact probability of V exceeding the observed value can be calculated as explained in §4 (*b*).

Example 3. Table 4 gives the analysis of dispersion for the three characters, head length (x_1), height (x_2) and weight (x_3) measured on 140 schoolboys, of almost the same age, belonging to six different schools in an Indian city.

Table 4. *Analysis of dispersion*

Dispersion	D.F.	S.P. matrix					
		x_1^2	x_2^2	x_3^2	x_1x_2	x_1x_3	x_2x_3
'Between' schools	5	(Q_{ij}) 752·0	151·3	1612·7	214·2	521·3	401·2
'Within' schools	134	(W_{ij}) 12809·3	1499·6	21009·6	1003·7	2671·2	4123·6
Total	139	(S_{ij}) 13561·3	1650·9	22622·3	1217·9	3192·5	4524·8

$$\Lambda = \frac{|W|}{|S|} = \frac{\begin{vmatrix} 12809\cdot3 & 214\cdot2 & 521\cdot3 \\ 214\cdot2 & 1499\cdot6 & 401\cdot2 \\ 521\cdot3 & 401\cdot2 & 21009\cdot6 \end{vmatrix}}{\begin{vmatrix} 13561\cdot3 & 1217\cdot9 & 3192\cdot5 \\ 1217\cdot9 & 1650\cdot9 & 4524\cdot8 \\ 3192\cdot5 & 4524\cdot8 & 22622\cdot3 \end{vmatrix}}$$

$$= \frac{10^{12}(0\cdot176005)}{10^{12}(0\cdot213628)},$$

$$-\log_e \Lambda = 0\cdot193724,$$
$$m = 139 - \tfrac{1}{2}(5+3+1) = 134\cdot5,$$
$$V = -m \log_e \Lambda = (134\cdot5)(0\cdot193724)$$
$$= 26\cdot0559.$$

Using V as χ^2 with $pq = 15$ degrees of freedom the first approximation comes out as

$$P_{15} = 0\cdot0375.$$

The second term is

$$\frac{\gamma_2}{m^2}(P_{19} - P_{15}),$$

where

$$\gamma_2 = \frac{29 \times 15}{48} \quad \text{and} \quad \frac{\gamma_2}{m^2} = \frac{29 \times 15}{48(134\cdot5)^2} = 0\cdot00050096,$$

$$\frac{\gamma_2}{m^2}(P_{19} - P_{15}) = 0\cdot00050096(0\cdot1285 - 0\cdot0375)$$
$$= 0\cdot00004574.$$

This correction to the first approximation affects only the fourth decimal place so that correction is hardly necessary. The observed value of V is significant at the 5 % level showing thereby that boys of various schools differ in physique. This appears to be generally true since boys belonging to different social strata attend different schools.

(d) *Internal analysis of a set of variates*

Let $x_1, \ldots, x_s, x_{s+1}, \ldots, x_{s+p}$ be $(s+p)$ correlated variables for which samples of sizes $N_1, \ldots, N_k$ are available from k populations. If the differences in mean values of these $(s+p)$ variables are to be tested for significance, then the method given in § 4 (c) can be used. An important problem that arises in biometry is to test whether the variables, say $x_{s+1}, \ldots, x_{s+p}$ bring out further differences in populations when the differences due to $x_1, \ldots, x_s$ are removed.

It is apparent in problems of this nature that some of the variables in the set $x_1, \ldots, x_s$ might be in the nature of concomitant variates which have been observed in association with the dependent variables or which might have been chosen to have some specified values. An illustration of such an analysis is given in my discussion on Bartlett's paper (Bartlett, 1947). In that problem there were three dependent variables g, h and i corresponding to linear, parabolic and cubic terms of growth curves of pig weights and a concomitant variable w giving the initial weight of pigs. It was desired to test whether the variables h and i bring out further differences in food treatments when the differences due to g and w are eliminated. The problem is identical with that posed above with g, w forming the first set and h, i the second set of variables.

There is a third set of problems in which it is desired to test whether the differences in k groups characterized by $(s+p)$ characters can be explained by variations in s *assigned* linear functions of these variables. If $y_1, \ldots, y_{s+p}$ are the $(s+p)$ variables and

$$L_1 = m_{1,1}y_1 + \ldots + m_{1,\,p+s}y_{p+s},$$

$$L_s = m_{s,1}y_1 + \ldots + m_{s,\,p+s}y_{p+s},$$

are the assigned linear functions, then one can replace the $(s+p)$ variables $y_1, \ldots, y_{s+p}$ by $x_1, \ldots, x_{s+p}$ defined by

$$x_1 = L_1, \quad \ldots, \quad x_s = L_s,$$

$$x_{s+1} = m_{s+1,1}y_1 + \ldots + m_{s+1,\,s+p}y_{s+p},$$

$$x_{s+p} = m_{s+p,1}y_1 + \ldots + m_{s+p,\,s+p}y_{s+p},$$

where the coefficients in $x_{s+1}, \ldots, x_{s+p}$ are chosen arbitrarily subject to the condition that the determinant $|m_{ij}|$, $i, j = 1, 2, \ldots, (s+p)$ is not zero. This latter condition ensures that the transformation from the y's to the x's leads to one-to-one correspondence. Once again, the problem is reduced to that of considering the differences in $x_{s+1}, \ldots, x_{s+p}$ when those due to $x_1, \ldots, x_s$ are removed. The proposed test is independent of the compounding coefficients used to define the set $x_{s+1}, \ldots, x_{s+p}$ so that, in any practical problem, they may be conveniently or conventionally chosen.

In all these cases, the problem is one of analysing the dispersion of the variables $x_{s+1}, \ldots, x_{s+p}$ when the dispersion due to $x_1, \ldots, x_s$ is removed. This can be done by following the covariance technique suitable for p dependent variables and s independent variables (Wishart, 1936, 1939; Rao, 1946a).

Let
$$(S_{ij}) = (Q_{ij}) + (W_{ij}) \quad (i, j = 1, 2, \ldots, (s+p))$$
be the analysis of dispersion for all the $(s+p)$ variates due to deviation from hypothesis and error with the corresponding distribution of degrees of freedom as
$$n' = q + (n' - q).$$
The S.P. matrix due to error for the variables $x_1, \ldots, x_s$ to be eliminated is
$$\begin{pmatrix} W_{11} & \cdots & W_{1s} \\ \cdots & \cdots & \cdots \\ W_{s1} & \cdots & W_{ss} \end{pmatrix}$$
and its inverse is represented by
$$\begin{pmatrix} W^{11} & \cdots & W^{1s} \\ \cdots & \cdots & \cdots \\ W^{s1} & \cdots & W^{ss} \end{pmatrix}.$$
The S.P. matrix due to error for $x_{s+1}, \ldots, x_{s+p}$ when corrected for $x_1, \ldots, x_s$, is given by $W(s+1, \ldots, s+p \mid 1, \ldots, s)$ or simply $W(p \mid s)$ where
$$W(p \mid s) = \begin{pmatrix} W_{s+1,s+1} & \cdots & W_{s+1,s+p} \\ \cdots & \cdots & \cdots \\ W_{s+p,s+1} & \cdots & W_{s+p,s+p} \end{pmatrix}$$
$$- \begin{pmatrix} W_{1,s+1} & \cdots & W_{s,s+1} \\ \cdots & \cdots & \cdots \\ W_{1,s+p} & \cdots & W_{s,s+p} \end{pmatrix} \begin{pmatrix} W^{11} & \cdots & W^{1s} \\ \cdots & \cdots & \cdots \\ W^{s1} & \cdots & W^{ss} \end{pmatrix} \begin{pmatrix} W_{1,s+1} & \cdots & W_{1,s+p} \\ \cdots & \cdots & \cdots \\ W_{s,s+1} & \cdots & W_{s,s+p} \end{pmatrix}.$$
This form which involves the evaluation of a triple product of matrices appears to be most convenient for computation as illustrated in the next section. Replacing W by S one has the formula for computing the S.P. matrix due to 'deviation from hypothesis + error' for $x_{s+1}, \ldots, x_{s+p}$ when corrected for $x_1, \ldots, x_s$. If this is represented by $S(p \mid s)$ then the required criterion is
$$\Lambda = \frac{|W(p \mid s)|}{|S(p \mid s)|}.$$
The degrees of freedom for $W(p \mid s)$ are $(n' - q - s)$ and for $S(p \mid s)$ are $(n' - s)$ so that in standard notation the parameters associated with Λ are
$$n = n' - s, \quad p = p, \quad q = q.$$
The test can be carried out as discussed in § 4 (b).

(e) *Barnard's problem of secular variations in skull characters*

The problem of measuring secular variations in skull characters considered by Barnard (1935) is of immense importance to the anthropologists. It is, however, of interest to examine the methods employed by her in the light of latest developments in multivariate analysis. The two problems involved in her study are

(i) the selection of a smaller number, out of seven skull characters, which give significant information, so far as is possible, as to changes taking place with time in four series of Egyptian skulls, and

(ii) the determination of an expression, linear in measurements, which characterizes most effectively an individual skull with respect to the progressive secular changes.

To answer problem (i) Barnard first chose Basi-alveolar Length and Nasal Height as two basic characters which independently of each other, show significant variation in the four

series. To choose further characters she considered the problem of testing the significance of the linear regression of the mean values of an added character with time (corresponding to the four series) when that part of the regression due to the two basic characters is removed. This meant the choice of characters with special reference to the average linear rate of change of the individual means with time. If the choice of characters is to be with reference to the complete nature of changes taking place with time, then what is needed is an internal analysis of the characters to decide whether the configuration of the four series as determined by several characters is the same as that indicated by a smaller number. Barnard's method should, of course, be preferred if the regressions were known to be linear. This can, however, be tested from the data.

Taking the four measurements

x_1 Basi-alveolar Length,
x_2 Nasal Height,
x_3 Maximum Breadth,
x_4 Basi-bregmatic Height,

the relevant data are summarized in Tables 5 and 6 which give the means for the four series and the analysis of dispersion.

Table 5. *Means for the four series*

Character	Series			
	I $N_1 = 91$	II $N_2 = 162$	III $N_3 = 70$	IV $N_4 = 75$
x_1	133·582418	134·265432	134·371429	135·306667
x_2	98·307692	95·462963	95·857143	95·040000
x_3	50·835165	51·148148	50·100000	52·093333
x_4	133·000000	134·882716	133·642857	131·466667

Table 6. *Analysis of dispersion*

	Dispersion		
	'Between' (3 D.F.)	'Within' (394 D.F.)	Total (397 D.F.)
x_1^2	123·180628	9661·997470	9785·178098
x_2^2	486·345863	9073·115027	9559·460890
x_3^2	100·411505	3938·320351	4088·731856
x_4^2	640·733891	8741·508829	9382·242720
x_1x_2	−231·375635	445·573301	214·197666
x_1x_3	87·305348	1130·623900	1217·929248
x_1x_4	−128·763994	2148·584210	2019·820216
x_2x_3	−107·505618	1239·221990	1131·716372
x_2x_4	125·313318	2255·812722	2381·126040
x_3x_4	−137·580764	1271·054662	1133·473898

Example 4. Do the characters x_3 and x_4 show significant variation in the four series independently of the variation due to the characters x_1 and x_2?

The method developed in § 4 (*d*) is directly useful in this problem. The S.P. matrix 'within' for the basic characters x_1 and x_2 is

$$\begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} = \begin{pmatrix} 9661{\cdot}997470 & 445{\cdot}573301 \\ 445{\cdot}573301 & 9073{\cdot}115027 \end{pmatrix}.$$

Its inverse is

$$\begin{pmatrix} W^{11} & W^{12} \\ W^{21} & W^{22} \end{pmatrix} = 10^{-4}\begin{pmatrix} 1{\cdot}037332 & -0{\cdot}050942 \\ -0{\cdot}050942 & 1{\cdot}104659 \end{pmatrix}.$$

The 'within' S.P. matrix for x_3, x_4 due to x_1, x_2 is given by the triple product

$$\begin{aligned}
&\begin{pmatrix} W_{13} & W_{23} \\ W_{14} & W_{24} \end{pmatrix}\begin{pmatrix} W^{11} & W^{12} \\ W^{21} & W^{22} \end{pmatrix}\begin{pmatrix} W_{13} & W_{14} \\ W_{23} & W_{24} \end{pmatrix} \\
&= \begin{pmatrix} 1130{\cdot}623900 & 1239{\cdot}221990 \\ 2148{\cdot}584210 & 2255{\cdot}812722 \end{pmatrix}\begin{pmatrix} W^{11} & W^{12} \\ W^{21} & W^{22} \end{pmatrix}\begin{pmatrix} W_{13} & W_{14} \\ W_{23} & W_{24} \end{pmatrix} \\
&= 10^{-4}\begin{pmatrix} 1109{\cdot}703904 & 1311{\cdot}321492 \\ 2113{\cdot}879535 & 2382{\cdot}450625 \end{pmatrix}\begin{pmatrix} W_{13} & W_{14} \\ W_{23} & W_{24} \end{pmatrix} \\
&= \begin{pmatrix} 287{\cdot}967620 & 534{\cdot}238796 \\ 534{\cdot}238796 & 991{\cdot}621041 \end{pmatrix}.
\end{aligned}$$

The 'within' S.P. matrix for x_3 and x_4 after correcting for x_1 and x_2 is

$$\begin{aligned}
&\begin{pmatrix} W_{33} & W_{34} \\ W_{34} & W_{44} \end{pmatrix} - \begin{pmatrix} W_{13} & W_{23} \\ W_{14} & W_{24} \end{pmatrix}\begin{pmatrix} W^{11} & W^{12} \\ W^{21} & W^{22} \end{pmatrix}\begin{pmatrix} W_{13} & W_{14} \\ W_{23} & W_{24} \end{pmatrix} \\
&= \begin{pmatrix} 3938{\cdot}320351 & 1271{\cdot}054662 \\ 1271{\cdot}054662 & 8741{\cdot}508829 \end{pmatrix} - \begin{pmatrix} 287{\cdot}967620 & 534{\cdot}238796 \\ 534{\cdot}238796 & 991{\cdot}621041 \end{pmatrix} \\
&= \begin{pmatrix} 3650{\cdot}353731 & 736{\cdot}815866 \\ 736{\cdot}815866 & 7749{\cdot}887788 \end{pmatrix} = W(2 \mid 2).
\end{aligned}$$

This has $394-2 = 392$ degrees of freedom. Similarly $S(2 \mid 2)$ with $397-2 = 395$ degrees of freedom is

$$\begin{pmatrix} 3809{\cdot}335190 & 611{\cdot}798381 \\ 611{\cdot}798381 & 8393{\cdot}755848 \end{pmatrix},$$

$$\Lambda = \frac{|W(2 \mid 2)|}{|S(2 \mid 2)|} = \frac{0{\cdot}27746934}{0{\cdot}31600332} = 0{\cdot}878058,$$

$$V = -m\log_e\Lambda, \quad m = n - \frac{p+q+1}{2} = 395 - \frac{2+3+1}{2} = 392,$$

$$V = -392\log_e(0{\cdot}878058) = 51{\cdot}39.$$

This value of V on $pq = 6$ degrees of freedom is highly significant so that x_3 and x_4 may be considered as discriminating the series independently of x_1 and x_2.

Example 5. Taking the relative times between the series in the proportion 2 : 1 : 2, can the variation of the characters be accounted for by the linear regression of individual characters with time?

In order to obtain the regression with time the values of t, the time variable, may be taken as -5, -1, $+1$ and $+5$ for the individuals of the first, second, third and fourth series respectively. The calculation of individual regressions involves the quantities

$$\Sigma(t-\bar{t})^2 = 4307{\cdot}66832,$$

$$\Sigma x_1(t-\bar{t}) = 718{\cdot}76286, \quad \Sigma x_3(t-\bar{t}) = -410{\cdot}10194,$$

$$\Sigma x_2(t-\bar{t}) = -1407{\cdot}26075, \quad \Sigma x_4(t-\bar{t}) = -733{\cdot}42758.$$

The matrix R with 1 degree of freedom giving the squares and products due to regression is given in Table 7.

Table 7. *Matrix R with 1 degree of freedom*

	x_1	x_2	x_3	x_4
x_1	119·930358	−234·810812	68·428235	−122·377258
x_2	−234·810812	459·734449	−133·975163	−149·601596
x_3	68·428235	−133·975163	39·042852	− 69·824358
x_4	−122·377258	−149·601596	− 69·824358	124·874099

In the above table

$$R_{11} = [\Sigma x_1(t-\bar{t})]^2/\Sigma(t-\bar{t})^2, \quad R_{12} = [\Sigma x_1(t-\bar{t})]\,[\Sigma x_2(t-\bar{t})]/\Sigma(t-\bar{t})^2,$$

and so on. With these results one may analyse the dispersion of which a typical product $(x_1 x_2)$ is chosen below for illustration.

Table 8. *Analysis of dispersion*

Due to	D.F.	S.P. matrix $(x_1 x_2)$
Regression	1	−234·810812 (R_{ij})
*Deviation from regression	2	3·435177 (Q_{ij})
Total ('between' series)	3	−231·375635 $(R_{ij}+Q_{ij})$
'Within' series	394	445·573301 (W_{ij})
Total	397	214·197666 (S_{ij})
Deviation from regression + 'within'	396	449·008478 $(Q_{ij}+W_{ij})$

* This quantity is obtained by subtraction. The complete matrix $(Q_{ij}+W_{ij})$ obtained by the above method is given in table 9.

Table 9. *Matrix $(Q_{ij}+W_{ij})$ with 396 degrees of freedom*

x_1	x_2	x_3	x_4
9665·247740	449·008478	1149·501013	2142·197474
449·008478	9099·726441	1265·691535	2231·524444
1149·501013	1265·691535	4049·689004	1203·298256
2142·197474	2231·524444	1203·298256	9257·368621

To test the hypothesis that the regressions are linear one has to compare W and $Q+W$.

$$\Lambda = \frac{|W|}{|Q+W|} = \frac{0{\cdot}24269054 \times 10^{12}}{0{\cdot}26873816 \times 10^{12}} = 0{\cdot}90307436,$$

$$V = -\left\{396 - \frac{2+4+1}{2}\right\} \log_e (0{\cdot}90307436)$$

40·02.

The χ^2 approximation has $p \times q = 2 \times 4 = 8$ degrees of freedom, since Q has 2 degrees of freedom and there are four variables. The result is significant so that the regressions cannot be considered linear.

This test can be extended to examine whether a parabolic regression with time can explain the differences in mean values. The matrix Q giving the deviation from regression has then 1 degree of freedom and R due to regression 2.

To determine the coefficients of a linear compound which characterizes most effectively the secular changes in progress, Barnard maximized the ratio of the square of unweighted regression of the compound with time. It is doubtful, as Bartlett (1947) points out, whether such a linear compound can be used to specify an individual skull most effectively with respect to progressive changes, since linear regression with time does not adequately explain all the differences in the four series.

The variance ratios obtained in this paper in the case of two samples can also be derived from a general regression analysis by considering certain pseudo-variates, which have constant values for members of the same sample, as dependent variables and the observed values as independent variables (Bartlett, 1939; Fisher, 1940; Brown, 1947). Such an approach does not seem to be possible in the case of a single sample. On the other hand, the statistics defined in (2·1) and (2·2) can be used in any situation where there are a number of linear hypotheses to be tested with the use of the estimated deviations and their independently estimated dispersion matrix. The distribution of the statistic in (2·2) has been derived by the author (Rao, 1946*b*) under these general conditions.

I wish to thank Dr Wishart for his helpful criticism during the preparation of this paper.

REFERENCES

Barnard, M. M. (1935). The secular variation of skull characters in four series of Egyptian skulls. *Ann. Eugen., Lond.*, **7**, 89.

Bartlett, M. S. (1934). The vector representation of a sample. *Proc. Camb. Phil. Soc.* **30**, 327.

Bartlett, M. S. (1938). Further aspects of the theory of multiple regression. *Proc. Camb. Phil. Soc.* **34**, 33.

Bartlett, M. S. (1939). The standard errors of discriminant function coefficients. *J. Roy. Statist. Soc. Suppl.* **6**, 169.

Bartlett, M. S. (1947). Multivariate analysis. *J. Roy. Statist. Soc. Suppl.* **9**, 76.

Bose, R. C. & Roy, S. N. (1938). The distribution of the studentized D^2-statistic. *Sankhyā*, **4**, 19.

Brown, G. W. (1947). Discriminant functions. *Ann. Math. Statist.* **18**, 514.

Fairfield Smith (1936). A discriminant function for plant selection. *Ann. Eugen., Lond.*, **7**, 240.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Ann. Eugen., Lond.*, **7**, 179.

Fisher, R. A. (1938). The statistical utilization of multiple measurements. *Ann. Eugen., Lond.*, **8**, 376.

Fisher, R. A. (1939). The sampling distribution of some statistics obtained from nonlinear regression. *Ann. Eugen., Lond.*, **9**, 238.

Fisher, R. A. (1940). The precision of the discriminant function. *Ann. Eugen., Lond.*, **10**, 422.

Hotelling, H. (1931). The generalization of Student's ratio. *Ann. Math. Statist.* **2**, 360.

Hotelling, H. (1935). The relation between two sets of variates. *Biometrika*, **28**, 321.

Hotelling, H. (1947). A generalized T measure of multivariate dispersion. *Ann. Math. Statist.* **18**, 298 (Abstract 2).

Hsu, P. (1939*a*). On the generalized analysis of variance. *Biometrika*, **31**, 221.

Hsu, P. (1939*b*). On the distribution of the roots of certain determinantal equations. *Ann. Eugen., Lond.*, **9**, 250.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proc. Nat. Inst. Sci.* (India), **12**, 49.

Martin, E. A. (1935). A study of the Egyptian series of mandibles, with special reference to mathematical methods of sexing. *Biometrika*, **28**, 149.

Nair, U. S. (1939). The application of moment functions in the study of distribution laws in statistics. *Biometrika*, **30**, 274.

Neyman, J. & Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, **20** A, 175.

Neyman, J. & Pearson, E. S. (1931). On the problem of k samples. *Bull. int. Acad. Cracovie*, A, p. 460.

Pearson, E. S. & Neyman, J. (1930). On the problem of two samples. *Bull. int. Acad. Cracovie*, A, p. 73.

Plackett, R. L. (1947). An exact test for the equality of variances. *Biometrika*, **34**, 311.

Rao, C. R. (1945). Generalization of Markoff's theorem and tests of linear hypothesis. *Sankhyā*, **7**, 9.

Rao, C. R. (1946*a*). On the linear combination of observations and the general theory of least squares. *Sankhyā*, **7**, 237.

Rao, C. R. (1946*b*). Tests with discriminant functions in multivariate analysis. *Sankhyā*, **7**, 407.

Rao, C. R. (1947). The problem of classification and distance between two populations. Letter to *Nature*, **159**, 30.

Roy, S. N. (1939). p-statistics or some generalizations in analysis of variance appropriate to multivariate problems. *Sankhyā*, **4**, 381.

Wald, A. & Brookner, R. J. (1941). On the distribution of Wilks' statistic for testing independence of several groups of variables. *Ann. Math. Statist.* **12**, 137.

Welch, B. L. (1939). Note on discriminant functions. *Biometrika*, **31**, 218.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, **24**, 471.

Wilks, S. S. (1935). On the independence of k sets of normally distributed statistical variables. *Econometrica*, **3**, 309.

Wishart, J. (1928). The generalized product moment distribution in samples from a normal multivariate population. *Biometrika*, **20** A, 32.

Wishart, J. (1936). Tests of significance in analysis of covariance. *J. Roy. Statist. Soc. Suppl.* **3**, 79.

Wishart, J. (1939). Statistical treatment of animal experiments. *J. Roy. Statist. Soc. Suppl.* **6**, 1.

ALTERNATIVE SYSTEMS IN THE ANALYSIS OF VARIANCE

By N. L. JOHNSON

1. It is the purpose of this paper to compare the fundamental theoretical set-ups implied by certain well-known systems of approach to the analysis of variance. Differences in interpretation in the different systems are discussed, and attention is drawn to some particularly simple results in the theory associated with one of the systems. No attempt is made to place the systems in any order of general preference. It is the author's opinion that each has its own sphere of application, while consideration of problems from the viewpoints of more than one of the systems will often prove enlightening.

2. The power functions of tests used in the analysis of variance have been considered by Hsu (1941), and by Tang (1938) who has given tables by means of which the power may be evaluated numerically in certain cases.

In the particular case of testing for differences between k groups, the theoretical set-up used by these authors is

$$x_{ij} = B_i + z_{ij} \quad (i = 1, \ldots, k;\ j = 1, \ldots, n_i), \tag{1}$$

where x_{ij} corresponds to the jth observation in the ith group. B_i is a constant representing the expected value in the ith group and the z_{ij}'s are independent random variables, each with zero expected value and standard deviation σ.

The appropriate analysis of variance is:

Source	Sum of squares	Degrees of freedom	Mean square
Between groups	$\sum_{i=1}^{k} n_i(\bar{x}_{i.}-\bar{x}_{..})^2$	$k-1$	$\frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{x}_{i.}-\bar{x}_{..})^2$
Within groups	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i.})^2$	$N-k$	$\frac{1}{N-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i.})^2$

where $N = \sum_{i=1}^{k} n_i$ is the total number of observations; $\bar{x}_{i.} = n_i^{-1}\sum_{j=1}^{n_i} x_{ij}$ is the mean observed value in the ith group; $\bar{x}_{..} = N^{-1}\sum_{i=1}^{k} n_i\bar{x}_{i.}$ is the mean of all the observed values. The expected value of the between groups mean square is $\sigma^2 + \frac{1}{k-1}\sum_{i=1}^{k} n_i(B_i-\bar{B})^2$. The expected value of the within groups mean square is σ^2. $\left(\bar{B} = N^{-1}\sum_{i=1}^{k} n_i B_i.\right)$ The ratio

(between groups mean square)/(within groups mean square)

is used to test whether there is any difference between the B_i's. If there is no such difference the expected values of the numerator and denominator of the above ratio are each equal to σ^2; otherwise the expected value of the numerator is greater than σ^2.

If it be assumed that each of the z_{ij}'s is normally distributed, the ratio of the mean squares will be referred to the F distribution with degrees of freedom $\nu_1 = k-1$, $\nu_2 = N-k$. Large observed values of the ratio are regarded as significant. A suitable upper significance limit of the appropriate F distribution may be used as a formal critical limit for the mean square ratio. Tang showed that, for such a test with alternative hypotheses specified by (1) above, the power (i.e. the chance of establishing significance when the B_i's are not all equal) depends

on k, N and, say, $\theta = (k-1)^{-1}\sum_{i=1}^{k} n_i(B_i - \bar{B})^2/\sigma^2$ only. The power function is somewhat complicated in its mathematical expression, but Tang's tables make it possible to determine the chance that a given ratio θ would be established as significant at either the 5 or 1 % level.

3. The alternative approach described below involves a modification in the theoretical set-up and leads to a very simple form of power function in a particular, but common, case. Although the modification may not always be justifiable, it will often provide a more accurate model than set-up (1), apart from the greater simplicity in the resulting analysis.

This alternative form of theoretical set-up is constructed by replacing (1) by

$$x_{ij} = A + z'_i + z_{ij}, \tag{2}$$

where A is a constant and the z'_i's are independent random variables, each with expected value zero and standard deviation σ'. The z'_i's and the z_{ij}'s are also mutually independent. In (2) the constants B_i of (1) are replaced by the random variables $A + z'_i$, and the hypothesis $B_1 = B_2 = \ldots = B_k$ is replaced by the hypothesis $\sigma' = 0$.

It is evident that (2) is a suitable set-up if it is possible to regard the groups as being chosen at random from a large assemblage of groups. If the groups are fixed, (1) will be preferable. Sub-sampling by batches from a randomly selected sample of batches is a typical example where (2) is suitable; comparison of a number of distinct treatments of a material is a typical example where (1) is preferable. Set-ups (1) and (2) represent extreme cases. In general it is likely that an intermediate set-up of form

$$x_{ij} = B_i + z'_i + z_{ij}$$

would be most appropriate. Daniels (1939) and Eisenhart (1947) have discussed, in some detail, the factors affecting the relevance of set-ups (1) and (2) in any particular problem.

However, the same analysis of variance is suitable in both extreme cases and so is likely to be suitable for any intermediate case. In fact, under the assumptions summarized in (2):

The expected value of the between groups mean square is $\sigma^2 + \sigma'^2\left(N^2 - \sum_{i=1}^{k} n_i^2\right)\Big/N(k-1)$.

The expected value of the within groups mean square is σ^2. The ratio

(between groups mean square)/(within groups mean square)

is again suitable for testing the hypothesis $\sigma' = 0$. If it be assumed that the z_{ij}'s are normally distributed, the F significance limits may be used as described in § 2.

4. If it now be assumed that the z'_i's, as well as the z_{ij}'s, are normally distributed, it is possible to obtain the power function in the case $n_1 = n_2 = \ldots = n_k = n$, say, in a particularly simple form. From (2) it follows that

$$\begin{aligned} \bar{x}_{i.} &= A + z'_i + \bar{z}_{i.} \quad \text{where} \quad \bar{z}_{i.} = n^{-1}\sum_{j=1}^{n} z_{ij} \\ &= A + u_i \qquad\quad\ \text{where} \quad u_i = z'_i + \bar{z}_{i.}. \end{aligned}$$

The u_i's are independent normal variables each with expected value zero and standard deviation $\sqrt{(n^{-1}\sigma^2 + \sigma'^2)}$. The between groups sum of squares is

$$\sum_{i=1}^{k} n(\bar{x}_{i.} - \bar{x}_{..})^2 = \sum_{i=1}^{k} n(u_i - \bar{u})^2 \quad \text{where} \quad \bar{u} = k^{-1}\sum_{i=1}^{k} u_i,$$

and hence is distributed as $\chi^2(\sigma^2 + n\sigma'^2)$ with degrees of freedom $\nu = k-1$. The within groups sum of squares is

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{i.})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n}(z_{ij} - \bar{z}_{i.})^2,$$

and is distributed independently as $\chi^2\sigma^2$ with $\nu_2 = k(n-1)$ degrees of freedom. Hence the mean square ratio used in the analysis of variance test is distributed as

$$F \times \left(\frac{\sigma^2 + n\sigma'^2}{\sigma^2}\right),$$

with $\nu_1 = k-1$, $\nu_2 = k(n-1)$.

If F_α, the upper 100α % point of the F distribution, be used as a formal critical limit, the probability of rejection of the hypothesis $\sigma' = 0$ is

$$\text{Pr.}\left\{F\left(\frac{\sigma^2 + n\sigma'^2}{\sigma^2}\right) > F_\alpha\right\} = \text{Pr.}\left\{F > F_\alpha\left(1 + \frac{n\sigma'^2}{\sigma^2}\right)^{-1}\right\}. \tag{3}$$

This is the power of the test with respect to the alternative hypothesis specified by (2) and a particular (non-zero) value of σ'. When $\sigma' = 0$ the probability of rejection is α. As σ' increases $F_\alpha(1 + n\sigma'^2/\sigma^2)^{-1}$ decreases and the power increases. The expression (3), considered as a function of σ', is the power function of the analysis of variance test in this case.

It will be noted that the power function is a function of $n\sigma'^2/\sigma^2$. This is analogous to the fact that with set-up (1) the power function is a function of $(n/(k-1))\sum_{i=1}^{k}(B_i - \bar{B})^2/\sigma^2$. σ'^2 and $\sum_{i=1}^{k}(B_i - \bar{B})^2/(k-1)$, indeed, fulfil similar roles in the two systems.

5. The calculation of the power function from (3) is straightforward. Since

$$p(F) = \frac{(k-1)^{\frac{1}{2}(k-1)}[k(n-1)]^{\frac{1}{2}k(n-1)}}{B(\frac{1}{2}[k-1], \frac{1}{2}k[n-1])} \frac{F^{\frac{1}{2}(k-3)}}{[k(n-1) + (k-1)F]^{\frac{1}{2}(nk-1)}},$$

it follows that $\text{Pr.}\{F > F_\alpha(1 + n\sigma'^2/\sigma^2)^{-1}\}$ can be expressed as the incomplete beta function ratio

$$I_\phi(\tfrac{1}{2}k[n-1], \tfrac{1}{2}[k-1]),$$

where $$\phi = k(n-1)/\{k(n-1) + (k-1)F_\alpha(1 + n\sigma'^2/\sigma^2)^{-1}\}.$$

When the number of groups is odd and not large it is possible to evaluate (3) as a simple explicit function $\beta(\sigma'/\sigma)$ of σ'/σ. For example if $k = 3$,

$$p(F) = K_n[\tfrac{3}{2}(n-1) + F]^{-\frac{1}{2}(3n-1)},$$

where $$K_n = [\tfrac{3}{2}(n-1)]^{\frac{1}{2}(3n-5)}.$$

Hence
$$\beta(\sigma'/\sigma) = \int_{F_\alpha(1+n\sigma'^2/\sigma^2)^{-1}}^{\infty} p(F)\,dF$$
$$= K'_n[\tfrac{3}{2}(n-1) + F_\alpha(1 + n\sigma'^2/\sigma^2)^{-1}]^{-\frac{3}{2}(n-1)}, \tag{4}$$

where $$K'_n = K_n/\tfrac{3}{2}(n-1) = [\tfrac{3}{2}(n-1)]^{\frac{1}{2}(3n-7)}.$$

Since
$$\int_{F_\alpha}^{\infty} p(F)\,dF = \alpha,$$
$$K'_n[\tfrac{3}{2}(n-1) + F_\alpha]^{-\frac{3}{2}(n-1)} = \alpha,$$

and so
$$\beta(\sigma'/\sigma) = \alpha\left[\frac{\frac{3}{2}(n-1) + F_\alpha}{\frac{3}{2}(n-1) + F_\alpha(1 + n\sigma'^2/\sigma^2)^{-1}}\right]^{\frac{3}{2}(n-1)}. \tag{4 bis}$$

6. In many cases the hypothesis $B_1 = B_2 = \ldots = B_k$, or the corresponding hypothesis in set-up (2), $\sigma' = 0$, is unduly stringent. That is to say, a test procedure is required which will lead to acceptance in a high proportion of cases provided there is not *too much* difference between the groups. It is natural to use as measure of the amount of difference between the

groups a parameter on which the power function depends. The parameter to be used would therefore be $\sum_{i=1}^{k} n_i(B_i - \bar{B})^2/(k-1)$ in conjunction with set-up (1), or σ' with set-up (2). The critical limit for the mean square ratio should then be chosen so that the chance of rejection is less than α_1, say, for any value of the parameter less than some specified amount.

It will also be desirable that rejection shall take place in a high proportion of cases—at least $1-\alpha_2$, say—when the groups differ by more than a certain amount. That is, the test should have at least a certain minimum sensitivity.

The requirements described above may be summed up formally as follows:

(*a*) The power function of the test must be less than α_1 for all values of the difference parameter less than some known amount.

(*b*) The power function must be greater than $1-\alpha_2$ for all values of the difference parameter greater than a second known amount.

This type of problem occurs in the theory of statistical quality control, the limiting values of the parameter corresponding to 'Producer's Tolerance' or 'Process Average' and 'Consumer's Tolerance' or 'Lot Tolerance' respectively.

Evidently, conditions (*a*) and (*b*) cannot both be satisfied unless the data available are sufficiently numerous. The amount of data required can be found from a study of the appropriate power functions. Tang's tables are useful in this connexion when set-up (1) is suitable. The analysis is much simpler for set-up (2) in the case we have considered in § 3 *et seq.* In the next section certain interesting results will be derived for this particular situation.

7. It will be convenient to define the difference between the groups by the parameter σ'/σ, instead of simply σ'. It may be noted that σ'/σ is the ratio of a parameter representing between-groups variability to one representing within-groups variability, and so can be regarded as a measure of *relative* between-groups variability. We shall require the analysis of variance test to be such that

(*a*) the probability of rejection is less than α_1 if $\sigma'/\sigma < \lambda_1$;

(*b*) the probability of rejection is greater than $1-\alpha_2$ if $\sigma'/\sigma > \lambda_2$.

In the limiting case when (*a*) is just satisfied, it is clear that the critical limit F_0 must be such that

$$\Pr.\{F(1+n\lambda_1^2) > F_0\} = \alpha_1.$$

Hence $F_0 = F_{\alpha_1}(1+n\lambda_1^2)$, F_{α_1} being the upper $100\alpha_1$ % point of the F distribution with $\nu_1 = k-1$, $\nu_2 = k(n-1)$.

If condition (*b*) is also to be satisfied we must have

$$\Pr.\{F(1+n\lambda_2^2) > F_0\} > 1-\alpha_2,$$

i.e.

$$\Pr.\{F(1+n\lambda_2^2) > F_{\alpha_1}(1+n\lambda_1^2)\} > 1-\alpha_2.$$

This means that

$$\frac{F_{\alpha_1}}{F_{1-\alpha_2}} < \frac{1+n\lambda_2^2}{1+n\lambda_1^2}, \tag{5}$$

$F_{1-\alpha_2}$ being the lower $100\alpha_2$ % point of the F distribution with $\nu_1 = k-1$, $\nu_2 = k(n-1)$. The choice of λ_1, λ_2, α_1 and α_2 will depend on practical requirements. Once these are decided upon, the number of groups, k, and the number of samples per group, n, should be chosen so that (5) is satisfied, if possible.

Imagine k fixed, and consider the effect of increasing n. As n increases the ratio $F_{\alpha_1}/F_{1-\alpha_2}$ decreases steadily, approaching the limit $\chi^2_{\alpha_1}/\chi^2_{1-\alpha_2}$ as n approaches infinity. ($\chi^2_{\alpha_1}$ and $\chi^2_{1-\alpha_2}$ represent the upper $100\alpha_1$ % and lower $100\alpha_2$ % points respectively of the χ^2 distribution

with $(k-1)$ degrees of freedom.) As n increases the ratio $(1+n\lambda_2^2)/(1+n\lambda_1^2)$ increases steadily from 1 to $(\lambda_2/\lambda_1)^2$. If $\chi^2_{\alpha_1}/\chi^2_{1-\alpha_2}$ is less than $(\lambda_2/\lambda_1)^2$ there will be a number n_0 such that if $n \geq n_0$ condition (5) is satisfied. If $\chi^2_{\alpha_1}/\chi^2_{1-\alpha_2}$ is greater than $(\lambda_2/\lambda_1)^2$, on the other hand, it will not be possible to satisfy (5), however large n may be. $\chi^2_{\alpha_1}/\chi^2_{1-\alpha_2}$ is a decreasing function of k, approaching 1 as k tends to infinity. There will therefore be a minimum number of groups, k_0, below which it is impossible to satisfy (5). A short table of such minimum values is given below.

$\alpha_1=\alpha_2=0\cdot05$		$\alpha_1=\alpha_2=0\cdot01$	
λ_2/λ_1	k_0	λ_2/λ_1	k_0
1·5	35	1·5	68
2	14	2	25
2·5	9	2·5	16
3	7	3	12

It may be noted that if it is reasonable to assume that σ is constant, λ_2/λ_1 is the ratio of the 'unacceptable' to the 'acceptable' limit of between-groups variability.

8. The alternative systems developed in §§ 2 and 3 are of use in the interpretation of the analysis of variance for data arranged according to a cross-classification. We shall consider only the simple case of a two-way cross-classification into l 'rows' and m 'columns', there being n observations in each cell. The symbol x_{ijt} will be used to denote the tth observation in the cell belonging to the ith row and the jth column. The analysis of variance table is:

Source	Sum of squares	Degrees of freedom
Between rows	$mn\sum_{i=1}^{l}(\bar{x}_{i..}-\bar{x}_{...})^2$	$l-1$
Between columns	$ln\sum_{j=1}^{m}(\bar{x}_{.j.}-\bar{x}_{...})^2$	$m-1$
Interaction	$n\sum_{i=1}^{l}\sum_{j=1}^{m}(\bar{x}_{ij.}-\bar{x}_{i..}-\bar{x}_{.j.}+\bar{x}_{...})^2$	$(l-1)(m-1)$
Within cells	$\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{t=1}^{n}(x_{ijt}-\bar{x}_{ij.})^2$	$lmn-1$

where
$$\bar{x}_{ij.}=n^{-1}\sum_{t=1}^{n}x_{ijt},\quad \bar{x}_{i..}=(mn)^{-1}\sum_{j=1}^{m}\sum_{t=1}^{n}x_{ijt},\quad \bar{x}_{.j.}=(ln)^{-1}\sum_{i=1}^{l}\sum_{t=1}^{n}x_{ijt},$$
$$\bar{x}_{...}=(lmn)^{-1}\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{t=1}^{n}x_{ijt}.$$

The question now arises—should the ratio of the between-rows mean square to the within-cells mean square be used to test for differences between rows, or should the ratio to the interaction mean square be used? (A similar question arises, of course, in the analysis between columns.) The answer to this question depends on which of the two set-ups (6) and (7), shown below is the more appropriate.

First, extending (1), we have
$$x_{ijt}=A+R_i+C_j+I_{ij}+z_{ijt}. \tag{6}$$

As before the z_{ijt}'s are independent random variables each with expected value zero and standard deviation σ. The parameter A is introduced to simplify the mathematics and represents the overall average level of the character measured. R_i represents the average departure from this level in the ith row; C_j represents the average departure in the jth column. Without loss of generality it may be assumed that

$$\sum_{i=1}^{l} R_i = \sum_{j=1}^{m} C_j = 0.$$

I_{ij} represents the interaction, or departure from the linear set-up, in the cell belonging to the ith row and the jth column. Without loss of generality it may be assumed that

$$\sum_{i=1}^{l} I_{ij} = 0 = \sum_{j=1}^{m} I_{ij}.$$

The expected values of the various mean squares in the analysis of variance table are, under the conditions summarized by (6):

	Mean square	Expected value
(i)	Between rows	$\sigma^2 + mn \sum_{i=1}^{l} R_i^2/(l-1)$
(ii)	Between columns	$\sigma^2 + ln \sum_{j=1}^{m} C_j^2/(m-1)$
(iii)	Interaction	$\sigma^2 + n \sum_{i=1}^{l} \sum_{j=1}^{m} I_{ij}^2/(l-1)(m-1)$
(iv)	Within cells	σ^2

It is clear, therefore, that the ratio of (i) to (iv) should be used to test the hypothesis $R_1 = R_2 = \ldots = R_l = 0$ when set-up (6) is valid. As in the case of set-up (1), Tang's tables can be used to calculate the power of the test to establish significance when there is a true effect. If the interaction mean square were used instead of the within-cells mean square we should risk reaching an inconclusive result when significance could have been established (i.e. there will be an increase in the second kind of error).

Now consider the alternative set-up

$$x_{ijt} = A + R_i + C_j + z'_{ij} + z_{ijt}, \tag{7}$$

formed by replacing the parameters I_{ij} by the random variables z'_{ij}. The z'_{ij}'s are assumed to be independent of each other and of the z_{ijt}'s, each having expected value zero and standard deviation σ'. Under these conditions the expected values of the mean squares in the analysis of variance table are:

	Mean square	Expected value
(i)	Between rows	$\sigma^2 + n\sigma'^2 + mn \sum_{i=1}^{l} R_i^2/(l-1)$
(ii)	Between columns	$\sigma^2 + n\sigma'^2 + ln \sum_{j=1}^{m} C_j^2/(m-1)$
(iii)	Interaction	$\sigma^2 + n\sigma'^2$
(iv)	Within cells	σ^2

It is apparent that if the ratio (i)/(iv) be used to test the hypothesis $R_1 = R_2 = \ldots = R_l = 0$ when set-up (7) is valid, the numerator will tend to be increased by the σ'^2 effect without any corresponding increase in the denominator. There would thus be a bias towards rejection of the null hypothesis (i.e. there will be an increase in the first kind of error). In this case, therefore, the ratio (i)/(iii) would be a preferable criterion.

In the case of a two-way classification, therefore, unlike the case of k groups, the two forms of theoretical set-up lead to different procedures in the analysis of variance. As before, the two systems, in this case (6) and (7), may be regarded as extreme cases. An intermediate set-up of the form
$$x_{ijt} = A + R_i + C_j + I_{ij} + z'_{ij} + z_{ijt}$$
would possibly be a truer reflexion of the practical position. This being so, it may be as well to consider both the ratio with the within-cells mean square and that with the interaction mean square in the denominator. Since the biases introduced by using the inexact tests are in opposite directions it follows that if the verdicts given by both ratios agree, confidence may be placed in the joint decision. Otherwise, closer consideration must be given to the conditions of collection of the data, to decide whether (6) or (7) is the more appropriate set-up.

Considerations similar to those developed above may be applied to more complicated problems in the analysis of variance. As a general rule, if an interaction is represented by a random variable, it is necessary to take it into account when testing interactions of lower order involving the same factors. Otherwise it is not necessary to do so.

9. Randomization theory provides yet a third system of theoretical set-ups. The essential difference between randomization theory and the systems already described is clearly illustrated by comparison of the theoretical set-ups they imply in the case of simple classification by groups.

It is possible to apply randomization theory to such a situation only if it is reasonable to suppose that any individual in the sample could have occurred in any one of the groups. A typical case arises when the effects of a number of treatments are to be investigated. The experimental material is divided into a number of groups and each group is assigned to a particular treatment.

It is supposed that the experimental arrangements are such that each possible arrangement of the N individuals into k groups of size $n_1, n_2, \ldots, n_k$ respectively is equally likely. The process of randomization in selecting the groups is an attempt to make practice consistent with the theory in this respect. The null hypothesis may then be expressed: 'the observed value of the character measured will be the same for any one individual, whatever be the group in which it is placed.' On the null hypothesis, therefore, there are only N possible observed values $u_1, u_2, \ldots, u_N$. Any observed set of values $x_{11}, \ldots, x_{kn_k}$ is simply a rearrangement of the u_r's. Since x_{ij} is equally likely to have any of the N values $u_1, u_2, \ldots, u_N$, the theoretical set-up on the null hypothesis may be written
$$x_{ij} = y_{ij}, \tag{8}$$
where y_{ij} is a discontinuous random variable with distribution function
$$\Pr.\{y_{ij} = u_r\} = 1/N \quad (r = 1, 2, \ldots, N).$$
It will be noted that all the y_{ij}'s have the same distribution, but they are not independent. If $y_{ij} = u_r$, then, in general, no other of the y_{ij}'s can take the value u_r.

It has been shown (Fisher, 1935; Welch, 1937, 1938; Pitman, 1937) that in the simple cases of classification by groups and cross-classification considered in this paper the distribution of the mean square ratio on the randomization theory may be replaced by the normal theory F distribution with little fear of serious error. In these cases, therefore, randomization of an experiment ensures that critical limits based on normal theory shall be approximately valid.

If we denote $\bar{u} = N^{-1}\sum_{r=1}^{N} u_r$ by A, (8) may be written
$$x_{ij} = A + y'_{ij},$$

where $y'_{ij} = y_{ij} - A$. The y'_{ij}'s are discontinuous, dependent random variables, each having the same distribution with expected value zero. Logically, alternative hypotheses, corresponding to the existence of differences between the groups, could be introduced according with either of the systems described earlier in this paper. Symbolically these would be expressed

$$x_{ij} = A + B_i + y'_{ij}, \tag{9}$$

$$x_{ij} = A + z'_i + y'_{ij}. \tag{10}$$

(9) corresponds to set-up (1), and (10) to (2), so far as the representation of group differences is concerned.

Set-up (1) may be written in the form

$$x_{ij} = A + B_i + z_{ij}. \tag{11}$$

(B_i has been replaced by $A + B_i$, but this is merely a matter of convenience, and does not affect the nature of the set-up.) We also recall that set-up (2) is

$$x_{ij} = A + z'_i + z_{ij}. \tag{12}$$

Studying set-ups (9) to (12) we notice that while the set-ups (1) and (2) differ in the nature of between-groups variation which they specify, randomization theory implies a modification in the distribution of the individual random residual variation, replacing the independent z_{ij}'s by the dependent y'_{ij}'s.

Although four different set-ups are shown above, it seems that (10) would be used but rarely. The use of constant parameters to represent group effects appears to be more consistent with the ideas of randomization theory.

REFERENCES

DANIELS, H. E. (1939). *J. Roy. Statist. Soc. Suppl.* **6**, 186.
EISENHART, C. (1947). *Biometrics*, **3**, 1.
FISHER, R. A. (1935). *The Design of Experiments.* London: Oliver and Boyd.
HSU, P. L. (1941). *Biometrika*, **32**, 62.
PITMAN, E. J. G. (1937). *Biometrika*, **29**, 322.
TANG, P. C. (1938). *Statist. Res. Mem.* **2**, 126.
WELCH, B. L. (1937). *Biometrika*, **29**, 21.
WELCH, B. L. (1938). *Biometrika*, **30**, 149.

AN EXAMINATION AND FURTHER DEVELOPMENT OF A FORMULA ARISING IN THE PROBLEM OF COMPARING TWO MEAN VALUES

By ALICE A. ASPIN, B.A., *Leeds University*

1. Introduction

In a recent paper B. L. Welch (1947) has developed, by a formal process, an expansion applicable to the problem of comparing two mean values. It is the purpose of the present paper (*a*) to extend this expansion to some further terms, (*b*) to investigate the numerical behaviour of the expansion in some particular cases and (*c*) to consider the comparative merits of a rearranged form of the expansion.

If we have two normal populations with true means α_1 and α_2 respectively and true variances σ_1^2 and σ_2^2; and if we have samples of sizes n_1 and n_2 drawn respectively from these populations, yielding sample means $\bar{x}_1$ and $\bar{x}_2$ and sample variances s_1^2 and s_2^2; then $(\bar{x}_1-\bar{x}_2)$ is distributed normally about $(\alpha_1-\alpha_2)$ with variance $(\sigma_1^2/n_1+\sigma_2^2/n_2)$, and s_1^2 and s_2^2 are distributed as $\chi_1^2\sigma_1^2/(n_1-1)$ and $\chi_2^2\sigma_2^2/(n_2-1)$ respectively. Problems of some importance are either to assess the significance of the observed difference $(\bar{x}_1-\bar{x}_2)$, or to calculate as a function of $(\bar{x}_1-\bar{x}_2)$, s_1^2 and s_2^2, limits within which the population difference $(\alpha_1-\alpha_2)$ may be said to lie with a given probability. In his theoretical discussion of this problem Welch has found it convenient to consider it as a particular case of the following more general problem which, formally, is no more difficult to solve.

Suppose η is any population parameter, estimated by an observed quantity y which is normally distributed with variance $\sigma_y^2 = \sum_{i=1}^{k}\lambda_i\sigma_i^2$. Suppose that, in addition, the data provide estimates s_i^2 of the unknown variances σ_i^2 $(i = 1, 2, \ldots, k)$, based on f_i degrees of freedom, and distributed as

$$p(s_i^2)\,ds_i^2 = \frac{1}{\Gamma(\frac{1}{2}f_i)}\left(\frac{f_i s_i^2}{2\sigma_i^2}\right)^{\frac{1}{2}f_i-1}\exp\left[-\frac{1}{2}\frac{f_i s_i^2}{\sigma_i^2}\right]d\left(\frac{f_i s_i^2}{2\sigma_i^2}\right), \tag{1}$$

where the s_i^2 $(i = 1, 2, \ldots, k)$ are all statistically independent of each other and of y. Let $h(s_1^2, s_2^2, \ldots, s_k^2, P)$ be that function of s^2 and P such that the probability is P that $(y-\eta)$ falls short of $h(s_1^2, s_2^2, \ldots, s_k^2, P)$. Then, if we can find $h(s_1^2, s_2^2, \ldots, s_k^2, P)$ in general, the particular application to the case of the comparison of means will follow by setting $k = 2$, $\eta = (\alpha_1-\alpha_2)$, $y = (\bar{x}_1-\bar{x}_2)$; $\lambda_1 = 1/n_1$ and $\lambda_2 = 1/n_2$.

Writing, for convenience, $h(s^2)$ for $h(s_1^2, s_2^2, \ldots, s_k^2, P)$, Welch has shown that the integral equation which $h(s^2)$ must satisfy can be expressed symbolically in the form

$$\Theta I\left\{\frac{h(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right\} = P. \tag{2}$$

The notation used here is that $I(v)$ stands for the integral, from $-\infty$ to v, of the unit normal probability function and

$$\begin{aligned}\Theta &= \exp\left[-\Sigma\sigma_i^2\partial_i^2 - \tfrac{1}{2}\Sigma f_i\log\left(1-\frac{2\sigma_i^2\partial_i^2}{f_i}\right)\right]\\ &= \exp\left[\Sigma\frac{\sigma_i^4\partial_i^2}{f_i} + \tfrac{4}{3}\Sigma\frac{\sigma_i^6\partial_i^3}{f_i^2} + \text{etc.}\right]\\ &= 1+\Sigma\frac{\sigma_i^4\partial_i^2}{f_i} + \left\{\tfrac{4}{3}\Sigma\frac{\sigma_i^6\partial_i^3}{f_i^2} + \frac{1}{2}\left(\Sigma\frac{\sigma_i^4\partial_i^2}{f_i}\right)^2\right\} + \text{etc.},\end{aligned} \tag{3}$$

where ∂_i^r implies repeated differentiation with respect to w_i and subsequent equation of all w_j to σ_j^2.

When the f_i are all large, the solution of (2) is $h_0(w) = \xi\sqrt{(\Sigma\lambda_i w_i)}$, where ξ is the normal deviate such that $I(\xi) = P$. More generally let us write

$$h(w) = h_0(w) + h_1(w) + h_2(w) + \text{etc.}, \tag{4}$$

where $h_1(w)$ includes terms of order $1/f_i$, $h_2(w)$ terms of order $1/f_i^2$, and so on. Further, suppose that we have calculated terms up to and including the order $1/f_i^r$ and wish to obtain an extra term $h_{r+1}(w)$. To the required order we may then write

$$\begin{aligned} I\left\{\frac{h(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right\} &= I\left\{\frac{h_0(w)+\ldots+h_r(w)+h_{r+1}(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right\} \\ &= I\left\{\frac{h_0(w)+\ldots+h_r(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right\} + \frac{h_{r+1}(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\, I'\left\{\frac{h_0(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right\}. \end{aligned} \tag{5}$$

When we operate on this with Θ the second term will, to the order involved, need to be treated only with the unit part of (3). Hence (2) will give

$$\Theta I\left\{\frac{h_0(w)+\ldots+h_r(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right\} + \frac{h_{r+1}(\sigma^2)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\, I'(\xi) = I(\xi). \tag{6}$$

This equation gives $h_{r+1}(\sigma^2)$ explicitly. Since $h_{r+1}(w)$ is the same function of w_i and $h_{r+1}(s^2)$ the same function of s_i^2 as $h_{r+1}(\sigma^2)$ is of σ_i^2, we have therefore an explicit method of deriving successive terms in an expansion for $h(s^2)$.

Pursuing the symbolic method a stage further we may expand $I\left\{\frac{h_0(w)+\ldots+h_r(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right\}$ formally in a Taylor series, thus

$$I\left\{\frac{h_0(w)+\ldots+h_r(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right\} = \exp\left[\left\{\frac{h_0(w)+\ldots+h_r(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}} - \xi\right\} D\right] I(v), \tag{7}$$

where D^r denotes repeated differentiation with respect to v and subsequent equation of v to ξ. The operation Θ in (6) may then be regarded as acting first on the exponential on the right-hand side of (7), producing an expression involving D which must in turn be operated on $I(v)$. Following out this procedure, Welch (1947, p. 31) has obtained the following expressions as far as order $1/f_i^2$:

$$h_0(s^2) = \xi\sqrt{(\Sigma\lambda_i s_i^2)}, \quad h_1(s^2) = \xi\sqrt{(\Sigma\lambda_i s_i^2)}\left\{\frac{(1+\xi^2)}{4}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i}\right)}{(\Sigma\lambda_i s_i^2)^2}\right\},$$

$$h_2(s^2) = \xi\sqrt{(\Sigma\lambda_i s_i^2)}\left[-\frac{(1+\xi^2)}{2}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i^2}\right)}{(\Sigma\lambda_i s_i^2)^2} + \frac{(3+5\xi^2+\xi^4)}{3}\frac{\left(\Sigma\frac{\lambda_i^3 s_i^6}{f_i^2}\right)}{(\Sigma\lambda_i s_i^2)^3} - \frac{(15+32\xi^2+9\xi^4)}{32}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i}\right)^2}{(\Sigma\lambda_i s_i^2)^4}\right]. \tag{8}$$

If we introduce the notation

$$V_{ru} = \frac{\left(\Sigma\frac{\lambda_i^r s_i^{2r}}{f_i^u}\right)}{(\Sigma\lambda_i s_i^2)^r} \tag{9}$$

equations (8) may be more compactly written as

$$h_0(s^2) = \xi\sqrt{(\Sigma\lambda_i s_i^2)}, \quad \frac{h_1(s^2)}{h_0(s^2)} = \tfrac{1}{4}(1+\xi^2)\,V_{21},$$

$$\frac{h_2(s^2)}{h_0(s^2)} = \left[-\tfrac{1}{2}(1+\xi^2)\,V_{22} + \tfrac{1}{3}(3+5\xi^2+\xi^4)\,V_{32} - \tfrac{1}{32}(15+32\xi^2+9\xi^4)\,V_{21}^2\right]. \tag{10}$$

2. The development of further terms

To find $h_3(s^2)$ we must first extend equation (3) to include terms of order $1/f_i^3$. This gives

$$\Theta = 1+\left(\Sigma\frac{\sigma_i^4\partial_i^2}{f_i}\right)+\left\{\frac{4}{3}\left(\Sigma\frac{\sigma_i^6\partial_i^3}{f_i^2}\right)+\frac{1}{2}\left(\Sigma\frac{\sigma_i^4\partial_i^2}{f_i}\right)^2\right\}+\left\{2\left(\Sigma\frac{\sigma_i^8\partial_i^4}{f_i^3}\right)+\frac{4}{3}\left(\Sigma\frac{\sigma_i^4\partial_i^2}{f_i}\right)\left(\Sigma\frac{\sigma_i^6\partial_i^3}{f_i^2}\right)+\frac{1}{6}\left(\Sigma\frac{\sigma_i^4\partial_i^2}{f_i}\right)^3\right\}. \tag{11}$$

This expression must operate on the exponential

$$\exp\left[\left\{\frac{h_0(w)+\ldots+h_2(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}-\xi\right\}D\right] = \exp\left[\left\{\frac{h_0(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}-\xi\right\}D\right]\exp\left[\left\{\frac{h_1(w)+h_2(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right\}D\right]$$
$$= \exp\left[\left\{\sqrt{\left(\frac{(\Sigma\lambda_i w_i)}{(\Sigma\lambda_i\sigma_i^2)}\right)}-1\right\}\xi D\right]\left[1+\frac{h_1(w)\,D}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}+\left\{\frac{h_2(w)\,D}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}+\frac{1}{2}\frac{h_1^2(w)\,D^2}{(\Sigma\lambda_i\sigma_i^2)}\right\}\right.$$
$$\left.+\left\{\frac{h_1(w)\,h_2(w)\,D^2}{(\Sigma\lambda_i\sigma_i^2)}+\frac{1}{6}\frac{h_1^3(w)\,D^3}{(\Sigma\lambda_i\sigma_i^2)^{\frac{3}{2}}}\right\}\right]. \tag{12}$$

$h_1(w)$ and $h_2(w)$ are already known and can be substituted in here. When this is done the successive differentiations required by (11) can be carried out and the first term on the left-hand side of (6) evaluated for the case $(r+1) = 3$. This equation then gives $h_3(\sigma^2)$. The algebra involved is heavy and the details will not be given here. The eventual result is found to be

$$\frac{h_3(s^2)}{h_0(s^2)} = [(1+\xi^2)\,V_{23}-2(3+5\xi^2+\xi^4)\,V_{33}+\tfrac{1}{8}(15+32\xi^2+9\xi^4)\,V_{22}V_{21}$$
$$+\tfrac{1}{8}(75+173\xi^2+63\xi^4+5\xi^6)\,V_{43}-\tfrac{1}{12}(105+298\xi^2+140\xi^4+15\xi^6)\,V_{32}V_{21}$$
$$+\tfrac{1}{384}(945+3169\xi^2+1811\xi^4+243\xi^6)\,V_{21}^3]. \tag{13}$$

The labour involved in calculating groups of terms of successive orders increases rapidly with r. This is due to the increasing number of differential operations introduced with each new term in the expansion of Θ and also to the rapidly increasing complexity of the expression on which Θ operates. Hence, even for $(r+1) = 3$, a very large number of differentiations with respect to the w_i have to be carried out before the final equation of all w_j to σ_j^2. In extending the work still further to the terms of order $1/f_i^4$, the present author found it necessary to cover over 100 pages with algebra, details of which need not now be given. The eventual result is found to be

$$\frac{h_4(s^2)}{h_0(s^2)} = [-2(1+\xi^2)\,V_{24}+\tfrac{28}{3}(3+5\xi^2+\xi^4)\,V_{34}$$
$$-\tfrac{1}{4}(15+32\xi^2+9\xi^4)\{V_{23}V_{21}+\tfrac{1}{2}V_{22}^2\}$$
$$-\tfrac{3}{2}(75+173\xi^2+63\xi^4+5\xi^6)\,V_{44}$$
$$+\tfrac{1}{2}(105+298\xi^2+140\xi^4+15\xi^6)\{\tfrac{1}{3}V_{22}V_{32}+V_{21}V_{33}\}$$
$$+\tfrac{1}{4}(15+33\xi^2+11\xi^4+\xi^6)\,V_{44}$$
$$+\tfrac{1}{5}(735+2170\xi^2+1126\xi^4+168\xi^6+7\xi^8)\,V_{54}$$
$$-\tfrac{1}{64}(945+3169\xi^2+1811\xi^4+243\xi^6)\,V_{42}V_{21}^2$$
$$-\tfrac{1}{18}(945+3354\xi^2+2166\xi^4+425\xi^6+25\xi^8)\,V_{32}^2$$
$$-\tfrac{1}{32}(4725+16586\xi^2+10514\xi^4+1974\xi^6+105\xi^8)\,V_{21}V_{43}$$
$$+\tfrac{1}{96}(10395+42429\xi^2+31938\xi^4+7335\xi^6+495\xi^8)\,V_{32}V_{21}^2$$
$$-\tfrac{1}{6144}(135135+626144\xi^2+542026\xi^4+145320\xi^6+11583\xi^8)\,V_{21}^4]. \tag{14}$$

3. CHECKS

As so much heavy algebra has been involved in reaching these results, greater confidence will be placed in them if some independent checks are available. One possibility is to try to find an expansion of some function of $h(s^2)$ rather than of $h(s^2)$ itself. It happens, indeed, that the square $h^2(s^2)$ can be developed in terms of successive orders in $1/f_i$ by a similar method to that used above without involving quite as much labour. In this development the $\boldsymbol{\Theta}$ of equation (2), instead of operating on a normal probability integral, operates on the integral of the distribution of the square of a unit normal deviate. The first approximation to $h^2(s^2)$ is $\xi^2(\Sigma\lambda_i s_i^2)$ so that, analogously to equation (12), we find that $\boldsymbol{\Theta}$ has to operate on the product of two factors of which the first is $\exp\left[\left\{\frac{(\Sigma\lambda_i w_i)}{(\Sigma\lambda_i \sigma_i^2)}-1\right\}\xi^2 D\right]$. Since this does not contain a square-root sign, some of the labour of operating with $\boldsymbol{\Theta}$ is lightened. A full check of equation (13) using this alternative method has been carried out.

A full independent check of the terms of order $1/f_i^4$ has not been similarly obtained, but a check which is almost as satisfactory is obtained by putting $k = 1$. We then find that (14) reduces to

$$\frac{h_4(s^2)}{h_0(s^2)} = \frac{(-945-1920\xi^2+1482\xi^4+776\xi^6+79\xi^8)}{360\times 4^4 f^4}. \tag{15}$$

This agrees, as it should, with the term in $1/f^4$ in the expansion of the straightforward 'Student' deviate given by R. A. Fisher (1941, p. 151). It is difficult to imagine algebraical mistakes which could have been made in reaching (14) without invalidating this agreement in the particular case, $k = 1$.

4. A REARRANGEMENT OF THE EXPANSION

It will be seen on inspection of equations (10), (13) and (14) that $h_1(s^2)$, $h_2(s^2)$, $h_3(s^2)$ and $h_4(s^2)$ all contain a term having $(1+\xi^2)$ as a factor. Grouping these terms together, the total contribution to $h(s^2)$ from this source is

$$\tfrac{1}{4}\{\xi\surd(\Sigma\lambda_i s_i^2)\}(1+\xi^2)\{V_{21}-2V_{22}+4V_{23}-8V_{24}\ldots\}. \tag{16}$$

Further, we have

$$\begin{aligned}(\Sigma\lambda_i s_i^2)^2\{V_{21}-2V_{22}+4V_{23}-8V_{24}\ldots\} &= \left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i}\right)-2\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i^2}\right)+4\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i^3}\right)-8\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i^4}\right)\ldots\\ &= \left\{\Sigma\frac{\lambda_i^2 s_i^4}{f_i}\left(1-\frac{2}{f_i}+\frac{4}{f_i^2}-\frac{8}{f_i^3}\ldots\right)\right\}\\ &= \left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i+2}\right).\end{aligned} \tag{17}$$

Hence (16) gives a contribution to $h(s^2)$ which we may denote by $h_1'(s^2)$, where

$$h_1'(s^2) = \xi\surd(\Sigma\lambda_i s_i^2)\,\frac{(1+\xi^2)}{4}\,\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i+2}\right)}{(\Sigma\lambda_i s_i^2)^2}. \tag{18}$$

Other terms may be combined in a similar manner. Indeed, if we write

$$W_u = \frac{\left(\Sigma\frac{\lambda_i^{u+1}s_i^{2(u+1)}}{(f_i+2)(f_i+4)\ldots(f_i+2u)}\right)}{(\Sigma\lambda_i s_i^2)^{u+1}}, \tag{19}$$

we can rearrange the expansion of $h(s^2)$ in the form

$$h(s^2) = h'_0(s^2) + h'_1(s^2) + h'_2(s^2) + \text{etc.}, \tag{20}$$

where now

$$h'_0(s^2) = \xi\sqrt{(\Sigma\lambda_i s_i^2)}, \quad \frac{h'_1(s^2)}{h'_0(s^2)} = \tfrac{1}{4}(1+\xi^2)\,W_1,$$

$$\frac{h'_2(s^2)}{h'_0(s^2)} = [\tfrac{1}{3}(3+5\xi^2+\xi^4)\,W_2 - \tfrac{1}{32}(15+32\xi^2+9\xi^4)\,W_1^2],$$

$$\frac{h'_3(s^2)}{h'_0(s^2)} = [\tfrac{1}{8}(75+173\xi^2+63\xi^4+5\xi^6)\,W_3$$
$$-\tfrac{1}{12}(105+298\xi^2+140\xi^4+15\xi^6)\,W_2W_1 + \tfrac{1}{384}(945+3169\xi^2+1811\xi^4+243\xi^6)\,W_1^3]. \tag{21}$$

The $h'_4(s^2)$ contribution does not come out completely in terms of the W_u's, but involves other expressions in addition. While this contribution cannot on this account be said to be anomalous, the best way of expressing it is not obvious and it is not proposed to enter into a discussion of it here.

Going only as far as terms of order $1/f_i^3$ it will be seen that, by using the W_u's, we have reduced the total number of terms in $h(s^2)$ from eleven to seven. Moreover, the saving in the number of these terms will be more marked as we proceed to higher orders of $1/f_i^r$. In certain circumstances, therefore, particularly if we are computing several probability levels simultaneously, it seems that there might be some gain in using equations (21). However, the question of convergence must be considered before a statement of this kind can properly be made.

5. Numerical investigation

This will be confined to the case $k = 2$ which includes, in particular, the problem of comparing two mean values. It will further be assumed that $f_1 = f_2 = f$, as would happen, for instance, in the mean value problem if the samples drawn from the two populations were equal. (The common f would then be one less than the common sample size.) We are not, however, assuming anything about the relative sizes of the unknown population variances σ_1^2 and σ_2^2.

We shall confine ourselves here to a single probability level $P = 95\,\%$, so that $\xi = 1{\cdot}64485$. Let us then write

$$H_j = 1 + \frac{\left(\sum\limits_{r=1}^{j} h_r(s^2)\right)}{\xi\sqrt{\left(\sum\limits_{i=1}^{k}\lambda_i s_i^2\right)}}, \quad H'_j = 1 + \frac{\left(\sum\limits_{r=1}^{j} h'_r(s^2)\right)}{\xi\sqrt{\left(\sum\limits_{i=1}^{k}\lambda_i s_i^2\right)}}, \tag{22}$$

so that H_j and H'_j are the successive approximations to $h(s^2)/\xi\sqrt{(\Sigma\lambda_i s_i^2)}$, according as the original expansion or the rearranged form is used. The H's depend on the observed data only through the ratio s_1^2/s_2^2 or, equivalently, the ratio $\lambda_1 s_1^2/(\lambda_1 s_1^2 + \lambda_2 s_2^2)$. Numerical values in Table 1 are given against this latter ratio for $f = 6$, 12 and 18. The argument of $\lambda_1 s_1^2/(\lambda_1 s_1^2 + \lambda_2 s_2^2)$ is by tenths from 1·0 to 0·5 (the values 0·4 to 0·0 follow by symmetry and are not therefore shown).

Considering first the expansion in powers of $1/f_i$, the degree of accuracy obtainable is seen from a comparison of H_3 and H_4. When $\lambda_1 s_1^2/(\lambda_1 s_1^2 + \lambda_2 s_2^2)$ equals 1 (or 0) we appear to have four decimal places accurate when $f = 6$, and five decimal places when $f = 12$ or more. In this extreme case $h(s^2)/\xi\sqrt{(\Sigma\lambda_i s_i^2)}$ is a simple 'Student' deviate divided by ξ. We already know, of course, that the series representation of such a quantity converges rapidly (e.g.

Fisher, 1941). In the final column of our Table 1 we give for comparison the values of t/ξ calculated from Mrs M. Merrington's table of the percentage points of 'Student's' t (Merrington, 1942).

In the remainder of Table 1 we have no similar independent source against which we can make any checks and can only be guided by the relative sizes of the H_j. It will be seen that, as we move towards the middle of the range of $\lambda_1 s_1^2/(\lambda_1 s_1^2+\lambda_2 s_2^2)$, the difference between H_3 and H_4 tends to increase. As a result it appears that the accuracy available at $\lambda_1 s_1^2/(\lambda_1 s_1^2+\lambda_2 s_2^2)$ equal to $\frac{1}{2}$ is two decimal places for $f=6$, three decimals for $f=12$ and four decimals for $f=18$. The accuracy improves towards each end of the range of $\lambda_1 s_1^2/(\lambda_1 s_1^2+\lambda_2 s_2^2)$.

Table 1. *Successive approximations to* $h(s^2)/\xi\sqrt{(\Sigma\lambda_i s_i^2)}$

($k=2$, $f_1=f_2=f$, $P=95\%$, $\xi=1{\cdot}64485$)

$\frac{\lambda_1 s_1^2}{\lambda_1 s_1^2+\lambda_2 s_2^2}$	$f=6$							
	H_1	H_2	H_3	H_4	H_1'	H_2'	H_3'	t_P/ξ
1·0	1·1544	1·1784	1·1812	1·1814	1·1158	1·1334	1·1266	1·1814
0·9	1·1266	1·1478	1·1524	1·1531	1·0949	1·1125	1·1101	—
0·8	1·1050	1·1176	1·1217	1·1248	1·0787	1·0926	1·0930	—
0·7	1·0895	1·0924	1·0917	1·0934	1·0672	1·0765	1·0771	—
0·6	1·0803	1·0761	1·0698	1·0668	1·0602	1·0659	1·0656	—
0·5	1·0772	1·0703	1·0616	1·0559	1·0579	1·0623	1·0615	—

$\frac{\lambda_1 s_1^2}{\lambda_1 s_1^2+\lambda_2 s_2^2}$	$f=12$							
	H_1	H_2	H_3	H_4	H_1'	H_2'	H_3'	t_P/ξ
1·0	1·07720	1·08319	1·08354	1·08355	1·06617	1·07496	1·07600	1·0836
0·9	1·06330	1·06861	1·06919	1·06923	1·05426	1·06221	1·06356	—
0·8	1·05249	1·05564	1·05617	1·05636	1·04500	1·05111	1·05242	—
0·7	1·04478	1·04552	1·04544	1·04555	1·03838	1·04253	1·04339	—
0·6	1·04014	1·03908	1·03829	1·03810	1·03441	1·03713	1·03747	—
0·5	1·03859	1·03687	1·03578	1·03542	1·03308	1·03528	1·03541	—

$\frac{\lambda_1 s_1^2}{\lambda_1 s_1^2+\lambda_2 s_2^2}$	$f=18$							
	H_1	H_2	H_3	H_4	H_1'	H_2'	H_3'	t_P/ξ
1·0	1·05147	1·05413	1·05423	1·05423	1·04632	1·05130	1·05219	1·0542
0·9	1·04220	1·04456	1·04473	1·04474	1·03799	1·04238	1·04324	—
0·8	1·03500	1·03640	1·03655	1·03659	1·03150	1·03485	1·03557	—
0·7	1·02985	1·03018	1·03015	1·03017	1·02687	1·02915	1·02961	—
0·6	1·02676	1·02629	1·02606	1·02602	1·02409	1·02561	1·02584	—
0·5	1·02573	1·02497	1·02465	1·02458	1·02316	1·02440	1·02451	—

Turning now to the rearranged form of the expansion we notice immediately that at the beginning of the table H'_3 does not compare favourably with H_3, since the value of H'_3 is not close to t/ξ.

As we leave this end of the table the numerical differences between H'_3 and H_3 decrease. It appears moreover, from an examination of the relative sizes of H'_1, H'_2 and H'_3, that in the vicinity of the value of $\lambda_1 s_1^2/(\lambda_1 s_1^2+\lambda_2 s_2^2)$ equal to $\frac{1}{2}$, the rearranged expansion converges more quickly than the original one. There does not, however, seem to be a sufficiently strong case for systematically abandoning the original expansion everywhere in favour of the rearranged form, even although the latter involves fewer terms.

The present numerical investigation is a preliminary one and other possible rearrangements of the expansion are being considered before any general suggestions for computing tables of $h(s^2)$ are finally made. It will, perhaps, best summarize the work to the point reached if

Table 2. *Values of* $h(s^2)/\sqrt{(\lambda_1 s_1^2+\lambda_2 s_2^2)}$ *and equivalent* F

($k = 2$; $P = 95\%$)

$\dfrac{\lambda_1 s_1^2}{\lambda_1 s_1^2+\lambda_2 s_2^2}$	$h(s^2)/\sqrt{(\lambda_1 s_1^2+\lambda_2 s_2^2)}$ for			Equivalent F for		
	$f = 6$	$f = 12$	$f = 18$	$f = 6$	$f = 12$	$f = 18$
1·0	1·9432	1·7823	1·7341	6	12	18
0·9	1·897	1·7587	1·7184	6·9	14·3	21·7
0·8	1·85	1·738	1·7050	8·3	17·4	26·3
0·7	1·80	1·720	1·6945	11	21	32
0·6	1·75	1·708	1·6877	15	25	37
0·5	1·74	1·703	1·6853	17	27	39

we present again compactly in Table 2 the final result, H_4, for all the cases considered. A slight modification may be made at this stage, by multiplying through by ξ so that the quantity tabled is now $h(s^2)/\sqrt{(\Sigma\lambda_i s_i^2)}$. The quantities in the first line of the table are then the 'Student' deviates for $f = 6$, 12 and 18. Only as many decimal places are given as appear to be justified by the behaviour of the successive terms in the expansion.

Although $h(s^2)/\sqrt{(\Sigma\lambda_i s_i^2)}$ cannot, beyond the first line of the table, be derived by noting that some quantity follows a simple 'Student' distribution, it is nevertheless of interest to consider what degrees of freedom F, say, a 'Student' deviate would have to possess if it were to have the same percentage points as the values of $h(s^2)/\sqrt{(\Sigma\lambda_i s_i^2)}$ given. Accordingly we have shown these equivalent F's in the last three columns. These values of F enable one, perhaps, to appreciate better the trend of the figures given in the earlier columns.

6. Application to the comparison of two means

The quantity $h(s^2)$ was defined in the first section to satisfy the relation

$$\Pr.[(y-\eta) < h(s^2)] = P, \tag{23}$$

$h(s^2)$ is, of course, a function of P as well as of $s_1^2, s_2^2, \ldots, s_k^2$ and should, perhaps, have been denoted by $h(s^2, P)$, but $h(s^2)$ was used throughout to save a little trouble in writing. The

dependence on P is understood. Turning now to the particular case of comparing mean values we must, as has already been noted, set

$$y = (\bar{x}_1 - \bar{x}_2), \quad \eta = (\alpha_1 - \alpha_2), \quad \lambda_1 = \frac{1}{n_1}, \quad \lambda_2 = \frac{1}{n_2}, \quad f_1 = (n_1 - 1), \quad f_2 = (n_2 - 2), \tag{24}$$

where n_1 and n_2 are the respective sample sizes. Then (23) becomes

$$\Pr.[\{(\bar{x}_1 - \bar{x}_2) - (\alpha_1 - \alpha_2)\} < h(s^2)] = P. \tag{25}$$

If we write

$$v = \frac{\{(\bar{x}_1 - \bar{x}_2) - (\alpha_1 - \alpha_2)\}}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}, \tag{26}$$

(25) becomes

$$\Pr.\left[v < \frac{h(s^2)}{\sqrt{(\lambda_1 s_1^2 + \lambda_2 s_2^2)}}\right] = P. \tag{27}$$

The numerator of v is normally distributed about zero with variance $(\sigma_1^2/n_1 + \sigma_2^2/n_2)$. The quantity under the square-root in the denominator of v is an unbiased estimate of $(\sigma_1^2/n_1 + \sigma_2^2/n_2)$ since s_1^2 and s_2^2 are respectively unbiased estimates of σ_1^2 and σ_2^2. The ratio v is, in general, to be distinguished from the quantity

$$u = \frac{\{(\bar{x}_1 - \bar{x}_2) - (\alpha_1 - \alpha_2)\}}{\sqrt{\left\{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right\}}}, \tag{28}$$

which would be the appropriate one to use if it could be assumed that $\sigma_1 = \sigma_2 = \sigma$. The ratio u would then be referred to the t-distribution with $(n_1 + n_2 - 2)$ degrees of freedom. If u is referred to the t-distribution when, in fact $\sigma_1 \neq \sigma_2$ we are liable to be led into error, as has been shown by Welch (1938). These errors are not serious if the sample sizes are equal ($n_1 = n_2 = n$) for then u and v are the same quantity. Such error as there is is then due to referring u (or v) to the 'Student' distribution with $2f = 2(n-1)$ degrees of freedom when, in fact, as we have seen, the appropriate percentage point depends to some extent on the observed ratio s_1/s_2. The critical value of v will be read off from a table like Table 2, entering with $f = (n-1)$ and with the ratio of s_1^2/n to $(s_1^2/n + s_2^2/n)$.

To take a particular example, suppose that $n_1 = n_2 = 7$ and quantities $\bar{x}_1$, $\bar{x}_2$, s_1 and s_2 are observed such that

$$\bar{x}_1 - \bar{x}_2 = 3{\cdot}7, \quad \sqrt{(s_1^2/7 + s_2^2/7)} = 1{\cdot}2, \quad \frac{s_1^2/7}{s_1^2/7 + s_2^2/7} = 0{\cdot}64. \tag{29}$$

Then entering Table 2 with $f = 6$ and the ratio 0·64, we have

$$h(s^2)/\sqrt{(\lambda_1 s_1^2 + \lambda_2 s_2^2)} = 1{\cdot}77.$$

If now we wished to test whether the data were consistent with the hypothesis that $\alpha_1 = \alpha_2$, we could compare the numerical value of the ratio $3{\cdot}7/1{\cdot}2 = 3{\cdot}1$ with the significance level 1·77; this, it exceeds greatly. On the other hand, if we wished to compute limits within which $(\alpha_1 - \alpha_2)$ lies with given probability, we would have from (27)

$$\Pr.[\{3{\cdot}7 - (\alpha_1 - \alpha_2)\}/1{\cdot}2 < 1{\cdot}77] = 0{\cdot}95. \tag{30}$$

Whence, on rearrangement of the inequality

$$\Pr.[(\alpha_1 - \alpha_2) > 3{\cdot}7 - 2{\cdot}12 = 1{\cdot}58] = 0{\cdot}95. \tag{31}$$

Owing to symmetry, the corresponding value of $h(s^2)/\sqrt{(\lambda_1 s_1^2+\lambda_2 s_2^2)}$ for $P = 5\%$ will be *minus* 1·77. Thus analogously to (30)

$$\Pr.\left[\frac{\{3\cdot 7-(\alpha_1-\alpha_2)\}}{1\cdot 2} < -1\cdot 77\right] = 0\cdot 05, \tag{32}$$

whence
$$\Pr.[(\alpha_1-\alpha_2) > 3\cdot 7+2\cdot 12 = 5\cdot 82] = 0\cdot 05. \tag{33}$$

Combining this with (31)

$$\Pr.[1\cdot 58 < (\alpha_1-\alpha_2) < 5\cdot 82] = 0\cdot 95-0\cdot 05 = 0\cdot 90. \tag{34}$$

In this equation the limits calculated for $(\alpha_1-\alpha_2)$ are not limits given to us before the samples are drawn, but depend on the observed sample statistics. The probability statement has the meaning usually attached to such statements when the method of inverse probability is not being used. At the earlier stage in equation (30) the same holds true. The numbers 3·7, 1·2 and 1·77 entering into this equation are all functions of the sample statistics and the probability statement is strictly about these functions and not about connexions between three pure numbers regarded as fixed and known to us before the samples were drawn. In principle the present procedure is exactly the same as that already familiar in those problems where the 'Student' t-distribution is applicable. The only difference in detail is that, in a 'Student' problem, the number 1·77 in (30) would be replaced by a tabular 'Student' deviate not dependent on the sample statistics observed. In the problem being considered in the present paper, we obtained the figure 1·77 by entering a table one of whose arguments was $\lambda_1 s_1^2/(\lambda_1 s_1^2+\lambda_2 s_2^2)$, so that the value obtained depended, although not very critically, on the observed ratio s_1/s_2. But the principle is really unaltered and the inversion of the inequality (30) proceeds in precisely the same manner as when we are considering a problem soluble directly by the straightforward 'Student' distribution.

In the present example and in the numerical work of the previous section we have confined ourselves to the case $k = 2$ and then to $f_1 = f_2 = f$. In the problem of comparing two means it is when $f_1 \neq f_2$ that the need for calculations of the kind considered here is most apparent for then, as Welch (1938) has pointed out, the u and v of equations (26) and (28) are not the same criterion and the error involved in the usual reference of u to the t-distribution can be more pronounced. We have considered the $f_1 = f_2$ case here first, merely to simplify calculations in an initial numerical investigation. It is proposed in some future work to discuss numerically the case where $f_1 \neq f_2$ and to compare the merits of different ways in which final tables may then be presented. Some consideration will also be given to the case where more than two population variances and their corresponding estimates are involved, i.e. the general case $k > 2$.

REFERENCES

Fisher, R. A. (1941). The asymptotic approach to Behrens' integral, with further tables for the d test of significance. *Ann. Eugen., Lond.*, **11**, 141–72.

Merrington, Maxine (1942). Table of percentage points of the t-distribution. *Biometrika*, **32**, 311.

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, **29**, 350–62.

Welch, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika*, **34**, 28–35.

ON THE POWER FUNCTION OF THE LONGEST RUN AS A TEST FOR RANDOMNESS IN A SEQUENCE OF ALTERNATIVES

BY G. BATEMAN, *University College, London*

1. INTRODUCTION

It has been suggested that the distribution of the longest run in a sequence of alternatives might be used with advantage in quality control work and allied subjects. Mosteller (1941) considered the case of runs above and below the median for a sample of even size and derived a formula for the probability of getting at least one run of a given length or greater when the number of elements of each of the two kinds is the same. Thus, if the alternatives in the sequence are E and $\bar{E}$, he considered explicitly only the case of $2n$ elements, n of which are E and n of which are $\bar{E}$. It is the purpose of this paper to deal with the slightly more general case of unequal numbers of elements of the two kinds and to consider (i) the distribution of the longest run under the hypothesis of randomness, and (ii) the power function when the alternative hypothesis is that of positive dependence in the sequence both for the simple Markoff chain and when the structure of dependence is more complex. In (ii) the conditional power function technique, as given by David (1947) for the distribution of groups, is used, and the two criteria, length of longest run and number of groups, are compared with respect to the same alternative hypotheses.

2. DISTRIBUTION OF THE LONGEST RUN

It will be assumed that there is a sequence of r elements, r_1 of which are E and r_2 of which are $\bar{E}$, where $r_1+r_2=r$. H_0, the hypothesis to be tested, will be that the elements of the sequence are in a random order, and the criterion used to carry out the test will be the length of the longest run of either E's or $\bar{E}$'s. The total number of sequences which can be formed from the r elements is ${}^rC_{r_1}$; this is the fundamental probability set. In order to pick out from this set the sub-set of sequences containing at least one greatest run of a given length, say g, it will be necessary to consider the partitions of r_1 and of r_2 having k as the greatest part, where $k=1, 2, \ldots, g$, and to find the number of ways in which they can be combined to form a sequence with at least one part equal to g and no part greater than g. This may be achieved most simply, perhaps, by considering the different ways in which such partitions of r_1 and r_2 form $2t$ or $2t+1$ groups, where $t=1, 2, \ldots, r_1-g+1$ for $r_1 \geqslant r_2$. There will be no loss of generality in assuming $r_1 \geqslant r_2$, and it will be understood in what follows that $r_1 \geqslant r_2$.

Let $f_i(t, k)$, where $i=1, 2$, denote the number of compositions* of r_i elements into t parts of which the greatest part contains k elements. This can be expressed in terms of binomial coefficients by using the result that the number of compositions of r_i into t parts, none of which exceeds s in magnitude, is the coefficient of x^{r_i} in the expansion of $(x+x^2+\ldots+x^s)^t$, i.e. of $x^t\left(\dfrac{1-x^s}{1-x}\right)^t$. This coefficient is $\sum\limits_{j=0}^{t}(-)^j\,{}^tC_j\,{}^{r_i-js-1}C_{t-1}$, and expressing it in terms of the above notation we have

$$\sum_{k\leqslant s} f_i(t, k) = \sum_{j=0}^{t}(-)^j\,{}^tC_j\,{}^{r_i-js-1}C_{t-1}. \tag{1}$$

* Compositions of a number are merely partitions of a number in which the order is taken into account.

Table 1. *Probability of obtaining at least one greatest run, of either kind of element, of given length, g*

(The probabilities are obtained by dividing the number tabled by the corresponding value of ${}^{r}C_{r_1}$)

r	r_1	r_2	${}^{r}C_{r_1}$	Values of g: 19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
10	9	1	10	.	.	.	.	.	.	.	.	.	.	2	2	2	2	2	.	.	.	.
	8	2	45	.	.	.	.	.	.	.	.	.	.	.	3	6	9	12	12	3	.	.
	7	3	120	.	.	.	.	.	.	.	.	.	.	.	.	4	12	24	40	36	4	.
	6	4	210	.	.	.	.	.	.	.	.	.	.	.	.	.	5	20	51	91	43	.
	5	5	252	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	48	110	82	2
11	10	1	11	.	.	.	.	.	.	.	.	.	2	2	2	2	2	1	.	.	.	.
	9	2	55	.	.	.	.	.	.	.	.	.	.	3	6	9	12	15	9	1	.	.
	8	3	165	.	.	.	.		.	.	.	.	.	.	4	12	24	40	54	30	1	.
	7	4	330	.	.	.	.	.	.	.	.	.	.	.	.	5	20	50	100	107	48	.
	6	5	462	.	.	.	.	.	.	.	.	.	.	.	.	.	6	33	106	204	112	1
12	11	1	12	.	.	.	.	.	.	.	.	2	2	2	2	2	2	.	.	.	.	.
	10	2	66	.	.	.	.	.	.	.	.	.	3	6	9	12	15	15	6	.	.	.
	9	3	220	.	.	.	.	.	.	.	.	.	.	4	12	24	40	60	60	20	.	.
	8	4	495	.	.	.	.	.	.	.	.	.	.	.	5	20	50	100	165	140	15	.
	7	5	792	.	.	.	.	.	.	.	.	.	.	.	.	6	30	92	220	330	114	.
	6	6	924	.	.	.	.	.	.	.	.	.	.	.	.	.	12	72	224	408	206	2
13	12	1	13	.	.	.	.	.	.	.	2	2	2	2	2	2	1	.	.	.	.	.
	11	2	78	.	.	.	.	.	.	.	.	3	6	9	12	15	18	12	3	.	.	.
	10	3	286	.	.	.	.	.	.	.	.	.	4	12	24	40	60	78	58	10	.	.
	9	4	715	.	.	.	.	.	.	.	.	.	.	5	20	50	100	175	230	130	5	.
	8	5	1287	.	.	.	.	.	.	.	.	.	.	.	6	30	90	211	410	453	87	.
	7	6	1716	.	.	.	.	.	.	.	.	.	.	.	.	7	46	177	466	735	284	1

Table 1. (*continued*)

(The probabilities are obtained by dividing the number tabled by the corresponding value of ${}^{r}C_{r_1}$)

r	r_1	r_2	${}^{r}C_{r_1}$	Values of g: 19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
14	13	1	14	.	.	.	.	.	.	2	2	2	2	2	2	2	.	.	.	.	.	.
	12	2	91	.	.	.	.	.	.	.	3	6	9	12	15	18	18	9	1	.	.	.
	11	3	364	.	.	.	.	.	.	.	.	4	12	24	40	60	84	88	48	4	.	.
	10	4	1001	.	.	.	.	.	.	.	.	.	5	20	50	100	175	270	280	100	1	.
	9	5	2002	.	.	.	.	.	.	.	.	.	.	6	30	90	210	420	668	528	50	.
	8	6	3003	.	.	.	.	.	.	.	.	.	.	.	7	42	150	418	919	1167	300	.
	7	7	3432	.	.	.	.	.	.	.	.	.	.	.	.	14	100	378	968	1454	516	2
15	14	1	15	.	.	.	.	.	2	2	2	2	2	2	2	1	.	.	.	.	.	.
	13	2	105	.	.	.	.	.	.	3	6	9	12	15	18	21	15	6	.	.	.	.
	12	3	455	.	.	.	.	.	.	.	4	12	24	40	60	84	106	90	34	1	.	.
	11	4	1365	.	.	.	.	.	.	.	.	5	20	50	100	175	280	370	300	65	.	.
	10	5	3003	.	.	.	.	.	.	.	.	.	6	30	90	210	420	741	960	525	21	.
	9	6	5005	.	.	.	.	.	.	.	.	.	.	7	42	147	394	900	1655	1614	246	.
	8	7	6435	.	.	.	.	.	.	.	.	.	.	.	8	61	270	846	1954	2576	719	1
20	19	1	20	2	2	2	2	2	2	2	2	2	2	.	.	.	.	.	.	.	.	.
	18	2	190	.	3	6	9	12	15	18	21	24	27	27	18	9	1	.	.	.	.	.
	17	3	1140	.	.	4	12	24	40	60	84	112	144	180	196	168	96	20	.	.	.	.
	16	4	4845	.	.	.	5	20	50	100	175	280	420	600	815	960	880	470	70	.	.	.
	15	5	15504	.	.	.	.	6	30	90	210	420	756	1260	1980	2880	3540	3086	1190	56	.	.
	14	6	38760	.	.	.	.	.	7	42	147	392	882	1764	3234	5523	8442	10192	7007	1127	1	.
	13	7	77520	.	.	.	.	.	.	8	56	224	672	1680	3696	7392	13516	20936	21672	7556	112	.
	12	8	125970	.	.	.	.	.	.	.	9	72	324	1080	2973	7180	15708	30124	42164	24945	1391	.
	11	9	167960	.	.	.	.	.	.	.	.	10	90	456	1736	5464	14784	34030	58406	47640	5344	.
	10	10	184756	.	.	.	.	.	.	.	.	.	20	208	1140	4464	13180	34570	64058	58290	8194	2

It follows immediately that

$$f_i(t,s) = \sum_{j=1}^{t} (-)^{j+1}\, {}^{t}C_j[{}^{r_i-j(s-1)-1}C_{t-1} - {}^{r_i-js-1}C_{t-1}].$$

If $N(2t, g \mid r_1r_2)$ denotes the number of sequences of $2t$ groups when at least one group contains g elements and no group contains more than g elements, then

$$N(2t, g \mid r_1r_2) = 2\left[f_1(t,g) \sum_{k\leqslant g} f_2(t,k) + f_2(t,g) \sum_{k\leqslant g-1} f_1(t,k)\right]. \tag{2}$$

The factor 2 is introduced to allow for the fact that the sequence may begin with E or with $\bar{E}$. In the same way it may be seen that

$$N(2t+1, g \mid r_1r_2) = \left[f_1(t+1,g) \sum_{k\leqslant g} f_2(t,k) + f_2(t,g) \sum_{k\leqslant g-1} f_1(t+1,k)\right]$$
$$+\left[f_1(t,g) \sum_{k\leqslant g} f_2(t+1,k) + f_2(t+1,g) \sum_{k\leqslant g-1} f_1(t,k)\right]. \tag{3}$$

As it will be required later, $\phi(t_1, t_2, g)$ will be used for the expression

$$f_1(t_1,g) \sum_{k\leqslant g} f_2(t_2,k) + f_2(t_2,g) \sum_{k\leqslant g-1} f_1(t_1,k) \quad \text{provided that} \quad |t_1 - t_2| \leqslant 1.$$

Thus $N(2t, g \mid r_1r_2) = 2\phi(t,t,g)$ and $N(2t+1, g \mid r_1r_2) = \phi(t+1,t,g) + \phi(t,t+1,g)$. The enumeration of the required subset is completed by summing $N(2t, g \mid r_1r_2)$ and $N(2t+1, g \mid r_1r_2)$ over all groups, i.e. from $t = 1$ to $t = r_1 - g + 1$. If the number in this subset be denoted by $N(g \mid r_1r_2)$, then

$$N(g \mid r_1r_2) = \sum_{t=1}^{r_1-g+1} (2\phi(t,t,g) + \phi(t+1,t,g) + \phi(t,t+1,g)). \tag{4}$$

Hence in a sequence of r elements, r_1 of which are E and r_2 of which are $\bar{E}$ where $r_1 + r_2 = r$ and $r_1 \geqslant r_2$, the probability that the longest run consists of g elements is

$$P\{g \mid r_1r_2\} = \frac{N(g \mid r_1r_2)}{{}^{r}C_{r_1}},$$

and the complete probability distribution of the length of the longest run, given r_1 and r_2, is obtained by letting g take all possible values. Values of the function (4) for $r = 10$ to 15, and $r = 20$ are given in Table 1; values of the functions (2) and (3) for $r_1 = 14$, $r_2 = 6$, illustrating the form of the correlation between T, the number of groups, and g, are given in Table 2.

If, as is frequently the case in statistical applications, we require only the probability of the longest run having a length greater than or equal to a given value, say g_0, it will be considerably simpler to calculate $\sum_{g\geqslant g_0} \phi(t_1, t_2, g)$ directly; for using result (1) it follows that

$$\sum_{g\geqslant g_0} \phi(t_1,t_2,g) = {}^{r_1-1}C_{t_1-1}\,{}^{r_2-1}C_{t_2-1} - \prod_{i=1}^{2}\left(\sum_{j=0}^{t_i} (-)^j\, {}^{t_i}C_j\, {}^{r_i-j(g_0-1)-1}C_{t_i-1}\right) \quad \text{for} \quad |t_1 - t_2| \leqslant 1$$
$$= 0 \qquad\qquad \text{for} \quad |t_1 - t_2| > 1,$$

and for brevity we shall write $\phi(t_1, t_2, g \geqslant g_0)$ for this expression.

The probability that the length of the greatest run, g, is greater than or equal to g_0 is immediate, for

$$P\{g \geqslant g_0 \mid r_1r_2\} = \sum_{g=g_0}^{r_1} P\{g \mid r_1r_2\}$$
$$\frac{1}{{}^{r}C_{r_1}}\left[\sum_{t=1}^{r_1-g_0+1} (2\phi(t,t,g\geqslant g_0) + \phi(t+1,t,g\geqslant g_0) + \phi(t,t+1,g\geqslant g_0))\right].$$

In the same way

$$P\{g \leqslant g_0 \mid r_1 r_2\} = \frac{1}{{}^{r}C_{r_1}}\left[\sum_{t=a}^{r_2}(2\phi(t,t,g \leqslant g_0)+\phi(t+1,t,g \leqslant g_0)+\phi(t,t+1,g \leqslant g_0))\right],$$

where

$$\phi(t_1,t_2,g \leqslant g_0) = \sum_{g \leqslant g_0} \phi(t_1,t_2,g) = \prod_{i=1}^{2}\left(\sum_{j=0}^{t_i}(-)^{j\,t_i}C_j{}^{r_i-jg_0-1}C_{t_i-1}\right) \quad \text{for } |t_1-t_2| \leqslant 1$$

$$= 0 \qquad \text{for } |t_1-t_2| > 1$$

and

$$a = \left[\frac{r_1+g_0-1}{g_0}\right].$$

Table 2. *The joint distribution of T and g for the case $r_1 = 14$, $r_2 = 6$*

(T = number of groups in the sequence)

T \ g	14	13	12	11	10	9	8	7	6	5	4	3	2	Total
2	2	.	.	.	.	.	.	.	.	.	.	.	.	2
3	5	2	2	2	2	2	2	1	.	.	.	.	.	18
4	.	20	20	20	20	20	20	10	.	.	.	.	.	130
5	.	20	35	50	65	80	95	100	60	15	.	.	.	520
6	.	.	60	120	180	240	300	360	240	60	.	.	.	1560
7	.	.	30	100	210	360	550	780	900	610	100	.	.	3640
8	.	.	.	80	240	480	800	1200	1560	1160	200	.	.	5720
9	.	.	.	20	110	320	700	1300	2140	2590	1350	50	.	8580
10	.	.	.	.	50	200	500	1000	1750	2300	1300	50	.	7150
11	.	.	.	.	5	50	200	550	1225	2255	2410	455	.	7150
12	.	.	.	.	.	12	60	180	420	810	912	180	.	2574
13	.	.	.	.	.	.	7	42	147	392	735	392	1	1716
Total	7	42	147	392	882	1764	3234	5523	8442	10192	7007	1127	1	38760

It may be noted that if longest runs of the E's only are considered then (2) and (3) reduce to $2f_1(t,g)\,{}^{r_2-1}C_{t-1}$ and $f_1(t+1,g)\,{}^{r_2-1}C_{t-1}+f_1(t,g)\,{}^{r_2-1}C_t$ respectively.

If g_E be used to denote the length of the longest run of E's in a sequence, then

$$N\{g_E = s \mid r_1 r_2\} = \sum_{t=1}^{r_1-s+1} f_1(t,s)\,[{}^{r_2-1}C_{t-2}+2\,{}^{r_2-1}C_{t-1}+{}^{r_2-1}C_t]$$

$$= \sum_{t=1}^{r_1-s+1} f_1(t,s)\,{}^{r_2+1}C_t,$$

and

$$N\{g_E \geqslant s \mid r_1 r_2\} = \sum_{t=1}^{r_1-s+1} {}^{r_2+1}C_t \sum_{j=1}^{t}(-)^{j+1}\,{}^tC_j\,{}^{r_1-j(s-1)-1}C_{t-1}.$$

Writing ${}^tC_j\,{}^{r_2+1}C_t = {}^{r_2+1}C_j\,{}^{r_2-j+1}C_{t-j}$, interchanging the order of summation and using the relation that $\sum_{i=0}^{m} {}^mC_{k+i}\,{}^nC_i = {}^{m+n}C_{k+n}$, we have that

$$P\{g_E \geqslant s \mid r_1 r_2\} = \frac{1}{{}^{r}C_{r_1}}\left[\sum_{j=1}^{[r_1/s]}(-)^{j+1}\,{}^{r_2+1}C_j\,{}^{r-js}C_{r_2}\right].$$

This result will be referred to in § 9.

3. Conditional power function for the simple Markoff chain

It is clear that the length of the longest run may be used as a criterion for testing randomness in a sequence in the same way as the number of groups or runs. If a longest run of length g_0 be observed, then, as shown in § 2, the probability of obtaining a run of length greater than

or equal to g_0, or of length less than or equal to g_0, can be found. If either of these probabilities is less than some arbitrarily assigned significance level, then we may reject H_0, the hypothesis of randomness. But the possible alternative hypotheses will be different if we judge on the upper tail (i.e. for g large) or the lower tail (i.e. for g small). When the observed g_0 is so large that the hypothesis H_0 is rejected, then a possible alternative hypothesis, H_1, might be that there is positive dependence of some kind in the sequence; that is to say the elements are not independent and given that E has occurred it becomes more likely that it will occur again. On the other hand if the observed g_0 is small, then a possible alternative hypothesis might be that there is negative dependence in the sequence.

Consider a sequence of possible events $e_1, e_2, \ldots, e_r$ which may or may not be independent; it is well known that

$$P\{e_1 e_2 \ldots e_r\} = \prod_{i=1}^{r} P\{e_i \mid e_1 e_2 \ldots e_{i-1}\}.$$

If the events are independent then the relation reduces to

$$P\{e_1 e_2 \ldots e_r\} = \prod_{i=1}^{r} P\{e_i\}.$$

This is the basis of H_0, the hypothesis of randomness.

If there is dependence as in a simple Markoff chain, each event will be dependent only on the event immediately preceding it, and we shall have

$$P\{e_1 e_2 \ldots e_r\} = \prod_{i=1}^{r} P\{e_i \mid e_{i-1}\}.$$

This is the basis of H_1, the possible alternative hypothesis to H_0, and we shall take into account only the case where the dependence is positive.

Now consider a sequence of r events composed of alternatives E or $\bar{E}$, and let E_i denote the happening of the event E at the ith trial. H_0 and H_1 will be as above with capital letters replacing small ones. Under the binomial hypothesis, i.e. under H_0, we write

$$P\{E_i\} = p \quad \text{and} \quad P\{\bar{E}_i\} = q \quad \text{for} \quad i = 1, 2, \ldots, r, \text{ where } p+q = 1.$$

and we shall have

$$P\{g \mid r_1 r_2 H_0\} = \frac{N(g \mid r_1 r_2)\, p^{r_1} q^{r_2}}{\sum_{\text{all } g} N(g \mid r_1 r_2)\, p^{r_1} q^{r_2}} = \frac{N(g \mid r_1 r_2)}{{}^{r}C_{r_1}}.$$

The hypothesis, H_0, is rejected if $g \geqslant g_\alpha$ where

$$P\{g \geqslant g_\alpha \mid r_1 r_2 H_0\} = \alpha, \tag{5}$$

and g_α is chosen so that α is as near as possible to the chosen significance level. Such values of g_α are shown in Table 3 for $r = 10, 15, 20$, and α in the neighbourhood of 0·05. This table could clearly be extended if desired.

As a likely alternative to H_0, we take H_1, and suppose that

$$P\{E_1\} = P, \qquad P\{\bar{E}_1\} = Q \quad \text{where} \quad P+Q = 1,$$
$$P\{E_i \mid E_{i-1}\} = P_1, \quad P\{\bar{E}_i \mid E_{i-1}\} = Q_1,$$
$$P\{E_i \mid \bar{E}_{i-1}\} = P_2, \quad P\{\bar{E}_i \mid \bar{E}_{i-1}\} = Q_2 \quad \text{where} \quad P_j + Q_j = 1 \text{ for } j = 1, 2.$$

The conditional power function $P\{g \mid r_1 r_2 H_1\}$ follows in a straightforward way by considering the partitions of r_1 and r_2 as was done to obtain the distribution of g under H_0. For each partition within a given number of $2t$ or $2t+1$ groups the multiplying probabilities are the same, for all that matters is the number of transitions from E to $\bar{E}$ and back again. Thus for a given

sequence of $2t$ groups beginning with E there are $2t-1$ transitions, t from E to $\bar{E}$ and $t-1$ from $\bar{E}$ to E and the remaining r_1-t and r_2-t are permanences of E and $\bar{E}$ respectively. The probability of obtaining a given sequence of $2t$ groups is equal to

$$PP_1^{r_1-t}P_2^{t-1}Q_1^tQ_2^{r_2-t}+QQ_1^{t-1}Q_2^{r_2-t}P_1^{r_1-t}P_2^t,$$

which may be written $$\left(\frac{P}{P_2}+\frac{Q}{Q_1}\right)\left(\frac{P_2Q_1}{Q_2P_1}\right)^t P_1^{r_1}Q_2^{r_2}.$$

In the same way the probability of obtaining a given sequence of $t+1$ groups of E and t of $\bar{E}$ is $\frac{P}{P_1}\left(\frac{P_2Q_1}{Q_2P_1}\right)^t P_1^{r_1}Q_2^{r_2}$, and of t groups of E and $t+1$ of $\bar{E}$ is $\frac{Q}{Q_2}\left(\frac{P_2Q_1}{Q_2P_1}\right)^t P_1^{r_1}Q_2^{r_2}$.

Table 3. *Values of g_α such that $P\{g \geq g_\alpha \mid r_1r_2\} = \alpha$, where α is as near as possible to* 0·05

(The value of α is given in brackets)

r_1 or r_2 \ r	10	15	20
2	8 (0·067)	13 (0·029)	17 (0·047)
3	7 (0·033)	11 (0·035)	15 (0·035)
4	6 (0·024)	9 (0·055)	12 (0·036)
5	5 (0·040)	8 (0·042)	11 (0·049)
6	—	7 (0·039)	10 (0·038)
7	—	6 (0·053)	9 (0·034)
8	—	—	8 (0·035)
9	—	—	7 (0·046)
10	—	—	7 (0·032)

The joint probability distribution of $2t$ and g is given by

$$P\{2t,g\mid r_1r_2H_1\}=\frac{\phi(t,t,g)\left(\frac{P}{P_2}+\frac{Q}{Q_1}\right)\left(\frac{P_2Q_1}{Q_2P_1}\right)^t}{\sum\limits_t\sum\limits_g\left(\frac{P_2Q_1}{Q_2P_1}\right)^t\left\{\phi(t,t,g)\left(\frac{P}{P_2}+\frac{Q}{Q_1}\right)+\phi(t+1,t,g)\frac{P}{P_1}+\phi(t,t+1,g)\frac{Q}{Q_2}\right\}}$$

$$=\frac{\phi(t,t,g)\left(\frac{P}{P_2}+\frac{Q}{Q_1}\right)\left(\frac{P_2Q_1}{Q_2P_1}\right)^t}{\sum\limits_{t=1}^{r_2}\left(\frac{P_2Q_1}{Q_2P_1}\right)^t\left\{{}^{r_1-1}C_{t-1}\,{}^{r_2-1}C_{t-1}\left(\frac{P}{P_2}+\frac{Q}{Q_1}\right)+{}^{r_1-1}C_t\,{}^{r_2-1}C_{t-1}\frac{P}{P_1}+{}^{r_1-1}C_{t-1}\,{}^{r_2-1}C_t\frac{Q}{Q_2}\right\}},$$

and similarly

$$P\{2t+1,g\mid r_1r_2H_1\}=\frac{\left(\phi(t+1,t,g)\frac{P}{P_1}+\phi(t,t+1,g)\frac{Q}{Q_2}\right)\left(\frac{P_2Q_1}{Q_2P_1}\right)^t}{\sum\limits_{t=1}^{r_2}\left(\frac{P_2Q_1}{Q_2P_1}\right)^t\left\{{}^{r_1-1}C_{t-1}\,{}^{r_2-1}C_{t-1}\left(\frac{P}{P_2}+\frac{Q}{Q_1}\right)+{}^{r_1-1}C_t\,{}^{r_2-1}C_{t-1}\frac{P}{P_1}+{}^{r_1-1}C_{t-1}\,{}^{r_2-1}C_t\frac{Q}{Q_2}\right\}}.$$

The probability distribution of g is obtained by summing over all t from $t = 1$ to $r_1 - g + 1$, that is to say

$$P\{g \mid r_1 r_2 H_1\} = \frac{\sum_{t=1}^{r_1-g+1} \left\{\phi(t,t,g)\left(\frac{P}{P_2}+\frac{Q}{Q_1}\right)+\phi(t+1,t,g)\frac{P}{P_1}+\phi(t,t+1,g)\frac{Q}{Q_2}\right\}\left(\frac{P_2Q_1}{Q_2P_1}\right)^t}{\sum_{t=1}^{r_2}\left(\frac{P_2Q_1}{Q_2P_1}\right)^t\left\{{}^{r_1-1}C_{t-1}\,{}^{r_2-1}C_{t-1}\left(\frac{P}{P_2}+\frac{Q}{Q_1}\right)+{}^{r_1-1}C_t\,{}^{r_2-1}C_{t-1}\frac{P}{P_1}+{}^{r_1-1}C_{t-1}\,{}^{r_2-1}C_t\frac{Q}{Q_2}\right\}}. \tag{6}$$

For the power of the test we require $P\{g \geq g_\alpha \mid r_1 r_2 H_1\}$, where g_α is given by (5). This probability is obtained by summing the left-hand side of (6) from $g = g_\alpha$ to r_1, and in practice it is simpler to sum first with respect to g and then with respect to t using the expression for $\sum_{g=g_\alpha}^{r_1} \phi(t_1, t_2, g)$ given in § 2.

4. Comparison of certain power curves for the criteria, length of longest run and number of groups

Throughout the computation of the power curves it has been assumed that

$$P = PP_1 + QP_2. \tag{7}$$

This condition is arrived at by using the relation $P\{E_i\} = P\{E_{i-1}E_i\} + P\{\bar{E}_{i-1}E_i\}$ on the assumption that $P\{E_i\} = P$ and $P\{\bar{E}_i\} = Q$ for all i; that is to say we are assuming that the probability of the event E occurring at the ith trial when nothing is known about the results of the preceding trials is independent of i. This in effect implies that the start of the sequence of observations is a randomly selected point in a longer sequence following the same law.

The power function for either criterion is given as a function of the three parameters P, P_1 and P_2, but, on application of (7), reduces to a function of two parameters. If we take these as P_1 and P_2, we note that P is a constant, k, on straight lines whose equations are $kP_1 + (1-k)P_2 = k$. The power functions for the number of groups criterion, as plotted by David for $P = 0{\cdot}5$, $0{\cdot}6$ and $0{\cdot}75$, are then sections of the power surface cutting the (P_1, P_2) plane in the straight lines (1), (2) and (3) shown in Fig. 1. As the alternative hypothesis under consideration is $P_1 > P_2$, the curves have been taken as starting at the central diagonal $P_1 = P_2 = P$, that is to say we are not interested in the power for $P_1 < P_2$. Fig. 1 shows contours of the power surface for the number of groups criterion when $r_1 = r_2 = 10$ and the significance level is $0{\cdot}05$. The probability of establishing a difference, using this significance level, when $P_1 > P_2$ is about $0{\cdot}22$, $0{\cdot}56$ and $0{\cdot}89$ respectively on the three contours shown.

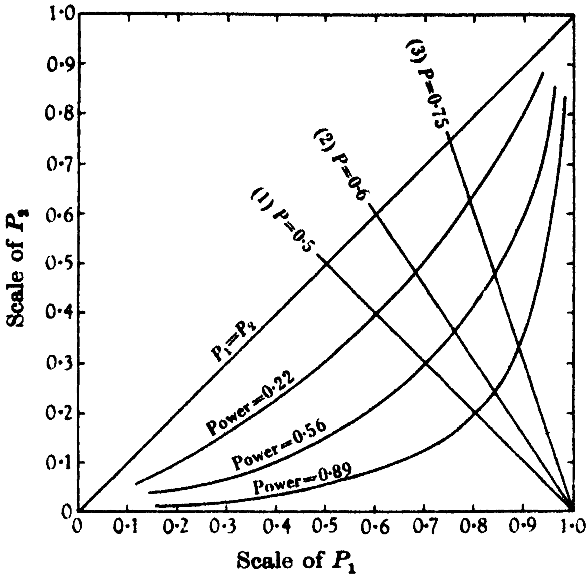
Alternatively we could express the power function in terms of P and $P_1 - P_2$. If, following Markoff (1913, p. 45), we take $P_1 - P_2 = \delta$, then $|\delta|$ is a measure of the degree of dependence in the sequence, and the sign of δ indicates the direction of dependence (+ for positive dependence, − for negative dependence). When $\delta = 0$, the observations are independent. Using (7) it follows that

$$P_1 = P + \delta Q, \quad Q_1 = Q(1-\delta), \quad P_2 = P(1-\delta), \quad Q_2 = Q + \delta P,$$

and the power function for either criterion may be expressed in terms of P and δ. The power curves for the two criteria when the alternative hypothesis is $\delta > 0$ are shown in Fig. 2 for $r_1 = r_2 = 10$, $r_1 = 14$, $r_2 = 6$ and $P = 0{\cdot}5$, $0{\cdot}6$, $0{\cdot}75$.* The chance, α, of rejecting H_0 when it is true is different for the two criteria; this is inevitable owing to the discontinuity of the dis-

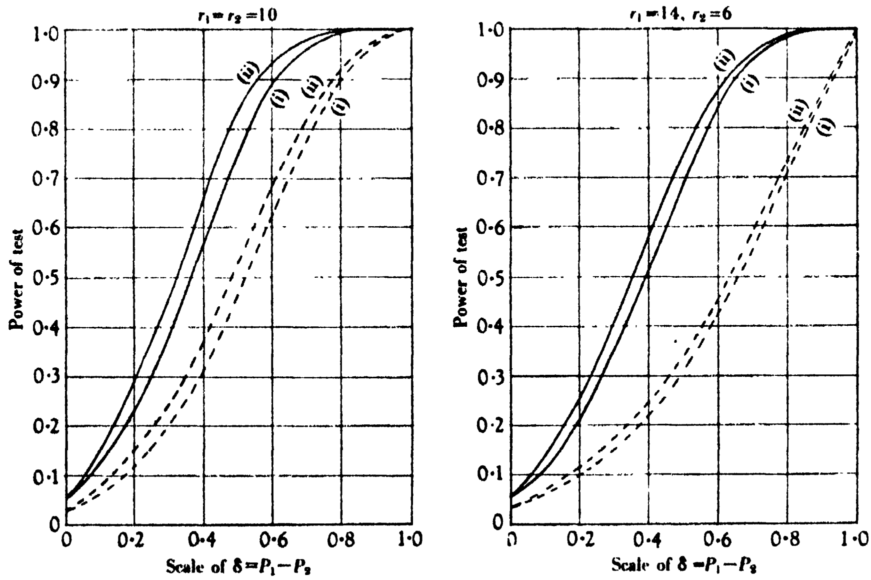
* When $P = 0{\cdot}5$ and $P = 0{\cdot}6$ the power curves are so close as to be indistinguishable on the graphs. This can be seen, too, from the contours in Fig. 1, for the contours are very nearly parallel to the diagonal $\delta = P_1 - P_2 = 0$, for this range of P.

tributions. For example, when $r_1 = 14$, $r_2 = 6$, $P\{g \geqslant 10 \mid H_0\} = 0{\cdot}038$ (see Table 3), while $P\{T \leqslant 6 \mid H_0\} = 0{\cdot}058$. These are the values of α nearest to the 0·05 level in each case. But, allowing for this difference, it is seen on comparing similar curves for the two criteria that T, the number of groups criterion, is the more powerful in detecting departures from randomness of the single dependence kind. This might perhaps have been expected on intuitive grounds.



The 5% significance level has been used.

Fig. 1. Contours of the power surface for the number of groups criterion, T. ($r_1 = r_2 = 10$)



Comparison of power curves for T and g.

Fig. 2. Case where the alternative hypothesis is positive dependence of the simple kind:

— $P\{T \leqslant T_\alpha \mid r_1 r_2 H_1\}$ --- $P\{g \geqslant g_\alpha \mid r_1 r_2 H_1\}$

For curves (i) $P = 0{\cdot}5$ or $0{\cdot}6$ (see footnote to § 4). For curves (ii) $P = 0{\cdot}75$.

It can further be noted that for either criterion the power is greater when $r_1 = r_2$, though the difference is more marked for the length of the longest run. In quality control work it is usual to consider runs above and below the median rather than above and below the mean. This ensures that $r_1 = r_2$ and gives increased power to the test.

5. The distribution of the number of groups when the hypothesis is that of double dependence in the chain

In § 4 it has been shown that T is a more powerful criterion than g for detecting departures from randomness when the alternative hypothesis is that of single dependence. It would be interesting to compare the powers of the tests when the alternative hypothesis is dependence of a different kind. Unfortunately the formulae become increasingly complex as the number of parameters defining the sequence is increased. We shall, therefore, deal only with the case of what could be called double dependence or dependence of the second order.

It is assumed in this case that each event is dependent only on the two events immediately preceding it. The general formula $P\{e_1 e_2 \ldots e_r\} = \prod_{i=1}^{r} P\{e_i \mid e_1 e_2 \ldots e_{i-1}\}$ then reduces to $P\{e_1 e_2 \ldots e_r\} = \prod_{i=1}^{r} P\{e_i \mid e_{i-2} e_{i-1}\}$. This is the basis of hypothesis H_2, and it is further assumed under H_2 that

(i) when nothing is known about the results of the $(i-2)$th and $(i-1)$th trials,

$$P\{E_i\} = P \quad \text{and} \quad P\{\bar{E}_i\} = Q, \quad \text{where} \quad P+Q = 1;$$

(ii) when the result of the $(i-1)$th trial is known, but not the result of the $(i-2)$th trial,

$$P\{E_i \mid E_{i-1}\} = P_1, \quad P\{\bar{E}_i \mid E_{i-1}\} = Q_1,$$
$$P\{E_i \mid \bar{E}_{i-1}\} = P_2, \quad P\{\bar{E}_i \mid \bar{E}_{i-1}\} = Q_2, \quad \text{where} \quad P_j + Q_j = 1 \text{ for } j = 1, 2;$$

(iii) when the result of the $(i-2)$th and $(i-1)$th trials are known,

$$P\{E_i \mid E_{i-2} E_{i-1}\} = p_1, \quad P\{\bar{E}_i \mid E_{i-2} E_{i-1}\} = q_1,$$
$$P\{E_i \mid E_{i-2} \bar{E}_{i-1}\} = p_2, \quad P\{\bar{E}_i \mid E_{i-2} \bar{E}_{i-1}\} = q_2,$$
$$P\{E_i \mid \bar{E}_{i-2} E_{i-1}\} = p_3, \quad P\{\bar{E}_i \mid \bar{E}_{i-2} E_{i-1}\} = q_3,$$
$$P\{E_i \mid \bar{E}_{i-2} \bar{E}_{i-1}\} = p_4, \quad P\{\bar{E}_i \mid \bar{E}_{i-2} \bar{E}_{i-1}\} = q_4, \quad \text{where} \quad p_j + q_j = 1 \text{ for } j = 1, 2, 3, 4.$$

Using the relations
$$P\{E_i\} = P\{E_{i-1} E_i\} + P\{\bar{E}_{i-1} E_i\},$$
$$P\{E_{i-1} E_i\} = P\{E_{i-2} E_{i-1} E_i\} + P\{\bar{E}_{i-2} E_{i-1} E_i\}$$
and
$$P\{\bar{E}_{i-1} E_i\} = P\{E_{i-2} \bar{E}_{i-1} E_i\} + P\{\bar{E}_{i-2} \bar{E}_{i-1} E_i\},$$
we obtain condition (7), i.e.
$$P = PP_1 + QP_2,$$
together with
$$P_1 p_1 + Q_1 p_3 = P_1, \tag{8}$$
and
$$P_2 p_2 + Q_2 p_4 = P_2. \tag{9}$$

These conditions again ensure that the start of the sequence is a point chosen at random in a longer sequence, and can be written in the form

$$P_1 = \frac{p_3}{1-p_1+p_3}, \quad P_2 = \frac{p_4}{1-p_2+p_4}, \quad P = \frac{P_2}{1-P_1+P_2} = \frac{p_4(1-p_1+p_3)}{p_4(1-p_1+p_3)+(1-p_1)(1-p_2+p_4)}.$$

Thus the problem is reduced to one with four parameters p_1, p_2, p_3, p_4.

To construct the power function under H_2 it is necessary to consider not only the sequences of $2t$ or $2t+1$ groups, but also the subset of such sequences containing a specified number of single E's and of single $\bar{E}$'s. If the r_1 E's are partitioned into t parts of which m_1 parts consist of a single E, then there are $t-m_1$ parts consisting of two or more E's, and the number of times that E follows EE is $r_1-m_1-2(t-m_1)$, i.e. r_1-2t+m_1. Similar results hold if the r_2 $\bar{E}$'s are partitioned into t parts of which m_2 parts consist of single $\bar{E}$'s. Suppose this sequence of $2t$ groups starts with at least two consecutive E's and ends with at least two consecutive $\bar{E}$'s. Then it can easily be seen that the probability of obtaining such a sequence is

$$PP_1q_1q_2p_1^{r_1-2t+m_1}q_4^{r_2-2t+m_2}(p_3q_1)^{t-m_1-1}(p_4q_2)^{t-m_2-1}p_2^{m_2}q_3^{m_1},$$

which we may write for brevity $\frac{PP_1}{p_3p_4}z$, where

$$z=p_1^{r_1}q_4^{r_2}\left(\frac{p_1q_3}{p_3q_1}\right)^{m_1}\left(\frac{p_2q_4}{p_4q_2}\right)^{m_2}\left(\frac{p_3p_4q_1q_2}{p_1^2q_4^2}\right)^t,$$

and refer to as sequence (i). The results for sequences (ii) starting EE, ending $E\bar{E}$, (iii) starting $E\bar{E}$, ending $\bar{E}\bar{E}$, (iv) starting $E\bar{E}$, ending $E\bar{E}$ are $\frac{PP_1}{p_2p_3}z$, $\frac{PQ_1}{p_4q_3}z$, $\frac{PQ_1}{p_2q_3}z$ respectively. Four corresponding results are obtained when the E's and $\bar{E}$'s in (i), (ii), (iii) and (iv) are interchanged.

For a sequence of $2t+1$ groups the subdivisions are into (a) $t+1$ groups of r_1 E's containing m_1 single E's and t groups of r_2 $\bar{E}$'s containing m_2 single $\bar{E}$'s, and (b) t groups of r_1 E's containing m_1 single E's and $t+1$ groups of r_2 $\bar{E}$'s containing m_2 single $\bar{E}$'s. Under (a) the sequences considered are those (i) starting EE, ending EE, (ii) starting EE, ending $\bar{E}E$, (iii) starting $E\bar{E}$, ending EE, (iv) starting $E\bar{E}$, ending $\bar{E}E$; and the probabilities of obtaining these sequences are $\frac{PP_1}{p_1^2}z$, $\frac{PP_1}{p_1^2}\frac{q_1}{q_3}z$, $\frac{PQ_1}{p_1^2}\frac{p_3}{q_3}z$, $\frac{PQ_1}{p_1^2}\frac{p_3q_1}{q_3^2}z$ respectively. Under (b) four corresponding subdivisions are considered.

Finally, the number of ways in which each such sequence can occur is required. The number of compositions of (r_1-m_1) E's into $(t-m_1)$ parts when no single E occurs is ${}^{r_1-t-1}C_{t-m_1-1}$ for $m_1=0,1,\ldots,t-1$, and is unity for $m_1=t=r_1$. This is the coefficient of $x^{r_1-m_1}$ in $(x^2+x^3+\ldots)^{t-m_1}$. Since the m_1 single E's can occupy the t spaces in ${}^tC_{m_1}$ ways, it follows that the number of compositions of r_1 E's into t parts, m_1 of which are single E's, is ${}^{r_1-t-1}C_{t-m_1-1}\cdot{}^tC_{m_1}$. If this be denoted by $h_1(t,m_1)$, then the number of sequences of $2t$ groups containing m_1 single E's and m_2 single $\bar{E}$'s is $2h_1(t,m_1)\,h_2(t,m_2)$ and the numbers of sequences of $2t$ groups specified by (i), (ii), (iii), (iv) are given by the product of $h_1(t,m_1)\,h_2(t,m_2)$ with

$$\frac{(t-m_1)(t-m_2)}{t^2},\quad \frac{(t-m_1)m_2}{t^2},\quad \frac{m_1(t-m_2)}{t^2},\quad \frac{m_1m_2}{t^2}$$

respectively. Similar results hold in the other cases.

Combining the results and summing for all m_1 and m_2, we obtain

$$P\{2t\,|\,r_1r_2H_2\}=\frac{A}{C}\quad\text{and}\quad P\{2t+1\,|\,r_1r_2H_2\}=\frac{B}{C}\quad\text{for}\quad t=1,2,\ldots,r_2\ \text{and}\ r_1\geqslant r_2,$$

where

$$\sum_{m_1=0}^{t}\sum_{m_2=0}^{t}\left(\frac{p_1q_3}{p_3q_1}\right)^{m_1}\left(\frac{p_2q_4}{p_4q_2}\right)^{m_2}\left(\frac{p_3p_4q_1q_2}{p_1^2q_4^2}\right)^t\frac{h_1(t,m_1)\,h_2(t,m_2)}{t^2}\left[(t-m_1)(t-m_2)\left(\frac{PP_1}{p_3p_4}+\frac{QQ_2}{q_1q_2}\right)\right.$$
$$\left.+(t-m_1)m_2\left(\frac{PP_1}{p_2p_3}+\frac{QP_2}{p_2q_1}\right)+m_1(t-m_2)\left(\frac{QQ_2}{q_2q_3}+\frac{PQ_1}{p_4q_3}\right)+m_1m_2\left(\frac{PQ_1+QP_2}{p_2q_3}\right)\right],$$

$$B = \sum_{m_1=0}^{t+1}\sum_{m_2=0}^{t+1}\left(\frac{p_1q_3}{p_3q_1}\right)^{m_1}\left(\frac{p_2q_4}{p_4q_2}\right)^{m_2}\left(\frac{p_3p_4q_1q_2}{p_1^2q_4^2}\right)^t$$

$$\times\left[\frac{h_1(t+1,m_1)\,h_2(t,m_2)}{t(t+1)}\frac{P}{p_1^2}\left\{(t-m_1+1)(t-m_1)P_1+m_1(t-m_1+1)\frac{P_1q_1+Q_1p_3}{q_3}\right.\right.$$

$$\left.+m_1(m_1-1)\frac{Q_1p_3q_1}{q_3^2}\right\}+\frac{h_1(t,m_1)\,h_2(t+1,m_2)}{t(t+1)}\frac{Q}{q_4^2}$$

$$\left.\times\left\{(t-m_2+1)(t-m_2)\,Q_2+m_2(t-m_2+1)\frac{Q_2p_4+P_2q_2}{p_2}+m_2(m_2-1)\frac{P_2q_2p_4}{p_2^2}\right\}\right]$$

$$C = \sum_{t=1}^{r_2}(A+B), \quad \text{and where} \quad h_i(t,m_i) = {}^{r_i-t-1}C_{t-m_i-1}\cdot{}^tC_{m_i} \quad \text{for} \quad m_i<t$$
$$= 1 \qquad\qquad \text{for} \quad m_i = t,\ i = 1,2.$$

Using conditions (7), (8), (9) these expressions simplify to some extent, and we may note:

(1) If $p_1 = p_2 = p_3 = p_4$, then $P = P_1 = P_2$ and $P\{T\mid r_1r_2H_2\}$ becomes $P\{T\mid r_1r_2H_0\}$, for A reduces to $2\,{}^{r_1-1}C_{t-1}{}^{r_2-1}C_{t-1}$ and B to ${}^{r_1-1}C_t{}^{r_2-1}C_{t-1}+{}^{r_1-1}C_{t-1}{}^{r_2-1}C_t$.

(2) If $p_1 = p_3$ and $p_2 = p_4$, then $P_1 = p_1$ and $P_2 = p_2$ and $P\{T\mid r_1r_2H_2\}$ becomes $P\{T\mid r_1r_2H_1\}$, for A reduces to $\left(\frac{P_2Q_1}{P_1Q_2}\right)^t\left(\frac{P}{P_2}+\frac{Q}{Q_1}\right){}^{r_1-1}C_{t-1}{}^{r_2-1}C_{t-1}$, and B reduces to

$$\left(\frac{P_2Q_1}{P_1Q_2}\right)^t\left[\frac{P}{P_1}{}^{r_1-1}C_t{}^{r_2-1}C_{t-1}+\frac{Q}{Q_2}{}^{r_1-1}C_{t-1}{}^{r_2-1}C_t\right].$$

(3) If $p_1 = p_2$ and $p_3 = p_4$, then $P = P_1 = P_2$ and we have what might be called 'throwback' dependence, for the result of the ith trial depends only on the result of the $(i-2)$th trial, and in fact the odd and even events in the sequence form independent sequences. The group test will be of little use unless we are interested in both positive and negative dependence as the alternative hypothesis.

(4) If $p_2 = p_3$, then the dependence is of the 'global' type, for each event depends only on the number of 'successes' in the preceding two trials.

6. The distribution of the length of the longest run under hypothesis H_2

This distribution can be obtained in a manner similar to the foregoing, but the formula is even more unwieldy. We shall consider, therefore, only $P\{g\geqslant s\mid r_1r_2H_2\}$. Suppose $h_i(t,m_i g<s)$, where $i = 1,2$, is the number of compositions of r_i elements into t parts of which m_i parts contain one and only one element and such that no part contains s or more than s elements. Then

$$h_i(t,m_i,g<s) = \left[\sum_{j=0}^{t-m_i}(-)^j\,{}^{t-m_i}C_j\,{}^{r_i-t-(s-2)j-1}C_{t-m_i-1}\right]{}^tC_{m_i} \quad \text{for} \quad s = 3,4,\ldots,r_i+1$$
$$= 1 \qquad\qquad \text{for} \quad s = 2,$$

where the expression in the square brackets is the coefficient of $x^{r_i-m_i}$ in the expansion of $(x^2+x^3+\ldots+x^{s-1})^{t-m_i}$. Clearly $h(t,m_i,g<r_i+1) = h(t,m_i)$. The number of sequences of $2t$ groups containing m_1 single E's and m_2 single $\bar{E}$'s and such that the longest run of either E's or $\bar{E}$'s has length greater than or equal to s is

$$2[h_1(t,m_1)\,h_2(t,m_2)-h_1(t,m_1,g<s)\,h_2(t,m_2,g<s)],$$

i.e. $$2\left[\prod_{i=1}^{2}{}^{r_i-t-1}C_{t-m_i-1}{}^tC_{m_i}-\prod_{i=1}^{2}\left(\sum_{j=0}^{t-m_i}(-)^j\,{}^{t-m_i}C_j\,{}^{r_i-t-(s-2)j-1}C_{t-m_i-1}{}^tC_{m_i}\right)\right].$$

If we denote by A' the expression obtained on replacing $h_1(t, m_1)\, h_2(t, m_2)$ in A (see § 5) by $h_1(t, m_1)\, h_2(t, m_2) - h_1(t, m_1, g < s)\, h_2(t, m_2, g < s)$, and if we obtain B' from B in a similar way, then

$$P\{g \geqslant s \mid r_1 r_2 H_2\} = \frac{\sum\limits_{t=1}^{r_1-s+1} (A'+B')}{\sum\limits_{t=1}^{r_2} (A+B)}.$$

7. Comparison of certain power curves for the criteria T and g when the alternative hypothesis is double dependence

In plotting the power curves, only the cases where

$$p_1 + p_4 = 1 = p_2 + p_3 \tag{10}$$

have been considered. Using the relations previously obtained, namely

$$P_1 = \frac{p_3}{1-p_1+p_3}, \quad P_2 = \frac{p_4}{1-p_2+p_4}, \quad P = \frac{P_2}{1-P_1+P_2},$$

it follows from (10) that $P_1 = Q_2$, $P_2 = Q_1$ and $P = 0{\cdot}5$. This seems a reasonable case to take for it merely implies symmetry with respect to E and $\bar{E}$. Writing the relations out in full we have $P = 0{\cdot}5$, $P_1 = Q_2$, $P_2 = Q_1$, $p_4 = q_1$, $p_1 = q_4$, $p_2 = q_3$ and $p_3 = q_2$. The power functions can now be expressed in terms of two parameters, say p_1 and p_3. Three sections of this power surface have been considered for $r_1 = r_2 = 10$, namely sections by the planes (*a*) $p_1 = p_3$ (H_1 in Fig. 2), (*b*) $p_3 = 0{\cdot}5$ (H_2' in Fig. 3), (*c*) $4p_3 - 2p_1 = 1$ (H_2'' in Fig. 3).

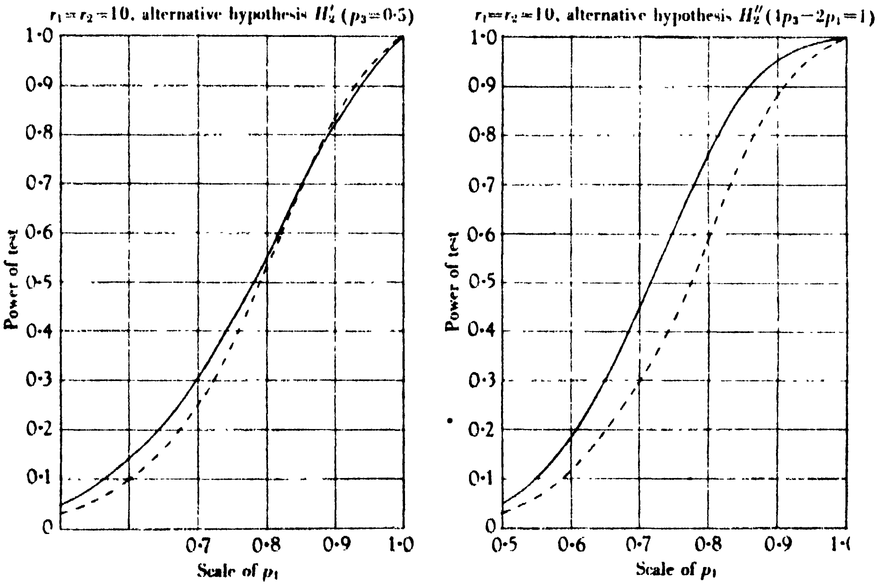


Fig. 3. Case where the alternative hypothesis is double dependence:
— $P\{T \leqslant T_\alpha \mid r_1 r_2 H_2\}$ - - - $P\{g \geqslant g_\alpha \mid r_1 r_2 H_2\}$

(*a*) When $p_1 = p_3$, as it does along the central diagonal in the (p_1, p_3) plane, the dependence in the sequence is of the single kind, and when $p_1 = p_3 = 0{\cdot}5$ then there is independence.

(*b*) When $p_3 = 0{\cdot}5$, then the dependence is of the global kind for the probability of obtaining a success at any trial depends on whether there were 2, 1, 0 successes in the preceding two trials; and in the case taken these probabilities decrease in arithmetic progression, e.g. when $p_1 = 0{\cdot}7$, $p_2 = p_3 = 0{\cdot}5$, $p_4 = 0{\cdot}3$.

(c) When $4p_3 - 2p_1 = 1$, the dependence is intermediate between these two types. Typical values of p_1, p_2, p_3, p_4 are shown in the following table:

p_1	0·5	0·6	0·7	0·8	0·9
p_3	0·5	0·55	0·6	0·65	0·7
p_2	0·5	0·45	0·4	0·35	0·3
p_4	0·5	0·4	0·3	0·2	0·1

It will be seen that, apart from the first column in which the events are independent, the probability of an event E occurring decreases according as the results of the two preceding trials are EE, $\bar{E}E$, $E\bar{E}$ or $\bar{E}\bar{E}$.

When the alternative hypothesis is single dependence, then, as we have noted previously, the g criterion is considerably less powerful than the T criterion in detecting departures from randomness. This point is illustrated by the power curves of Fig. 2. The power curves of Fig. 3 show that this is not necessarily the case when the alternative hypothesis is double dependence, for in one of the two cases considered the g criterion would appear to be no less powerful than T in detecting departures from the basic hypothesis.

It is realized that this investigation is not yet complete, but further work is being done on types of dependence likely to occur in practice, and on the possibility of using an estimate of the degree of dependence in a sequence as a test criterion.

8. Illustration of dependence in a sequence

As an illustration of sequences which might show dependence, two passages have been taken at random, the first from a standard text-book on Statistics and the second from Gertrude Stein's writings. The event, E, recorded is the occurrence of a word of one syllable, while $\bar{E}$ is the occurrence of a word of more than one syllable. To eliminate possible end effects the

First sample

Present word

Preceding word	E	$\bar{E}$	Total
E	194	133	327
$\bar{E}$	132	41	173
Total	326	174	500

Present word

Preceding two words	E	$\bar{E}$	Total
EE	108	85	193
$E\bar{E}$	101	32	133
$\bar{E}E$	86	48	134
$\bar{E}\bar{E}$	31	9	40
Total	326	174	500

Second sample

Present word

Preceding word	E	$\bar{E}$	Total
E	301	89	390
$\bar{E}$	89	21	110
Total	390	110	500

Present word

Preceding two words	E	$\bar{E}$	Total
EE	231	72	303
$E\bar{E}$	69	19	88
$\bar{E}E$	70	17	87
$\bar{E}\bar{E}$	20	2	22
Total	390	110	500

first three and the last three words in each sentence have not been counted. The results for the individual sentences have been pooled to obtain the above frequency tables. The total number of words counted was 500 in each case.

Estimates of the probabilities, relative probabilities and degree of dependence (δ) are shown in the following table:

	$\hat{P}$	$\hat{P}_1$	$\hat{P}_2$	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	$\hat{p}_4$	$\hat{\delta}$
1st sample	0·65	0·59	0·76	0·56	0·76	0·64	0·78	−0·17
2nd sample	0·78	0·77	0·81	0·76	0·78	0·80	0·91	−0·04

If we take as our hypothesis that $P_1 = P_2 = P$, then, applying the χ^2-test for independence to the 2×2 tables above, we obtain values for χ^2 of 13·62 for the first sample and 0·49 for the second sample. On this evidence we reject the hypothesis of independence in the first case, but not in the second case. The first sample does, in fact, seem to exhibit a degree of negative dependence of the single kind, i.e. given a monosyllabic (or polysyllabic) word the chance of its being followed by another monosyllabic (or polysyllabic) word is less than it would be if the events were independent. It is not clear whether or not the sequence shows double dependence. Applying the χ^2-test to the 2×4 table we again get a significant value for χ^2, which leads us to reject the hypothesis, $p_1 = p_2 = p_3 = p_4 = P$; but it might still be the case that $p_1 = p_3 = P_1$ and $p_2 = p_4 = P_2$, when the double dependence would reduce to single dependence.

In the second sample it might well be that $P = P_1 = P_2 = p_1 = p_2 = p_3 = p_4$, the fluctuations in the estimates being due to chance, and the χ^2-test applied to the 2×2 and to the 2×4 table confirms this. Thus in Gertrude Stein's work monosyllabic and polysyllabic words would seem to occur at random.

9. Application of the distribution of the longest run to the classical problem of 'runs of luck'

The problem of 'runs of luck' was first put forward and solved by de Moivre. It is concerned with finding the probability that an event E occurs at least s times in succession in a series of r independent trials, when the probability that the event E occurs is constant and equal to p. The problem has been solved using difference equations. The alternative solution given below is based on the distribution of the longest run.

It has been shown in § 2 that, if g_E is the length of the longest run of E's, then

$$P\{g_E \geqslant s \mid r_1 r_2\} = \sum_{j=1}^{[r_1/s]} (-)^{j+1}\, {}^{r_2+1}C_j\, {}^{r-js}C_{r_2} / {}^{r}C_{r_1}, \quad \text{where} \quad r_1 + r_2 = r.$$

If r (the number of independent trials) is given and $P\{E\}$ is constant and equal to p, then

$$P\{r_1\} = {}^{r}C_{r_1}\, p^{r_1} q^{r_2}$$

and

$$\begin{aligned} P\{g_E \geqslant s\} &= \sum_{r_1=s}^{r} P\{g_E \geqslant s \mid r_1\} \times P\{r_1\} \\ &= \sum_{r_1=s}^{r} \sum_{j \geqslant 1} (-)^{j+1}\, {}^{r_2+1}C_j\, {}^{r-js}C_{r_2}\, p^{r_1} q^{r_2} \\ &= \sum_{j \geqslant 1} (-)^{j+1}\, {}^{r-js}C_{j-1}\, p^{js} q^{j-1} \sum_{r_1 \geqslant js} \frac{r_2+1}{j}\, {}^{r-j(s+1)+1}C_{r_1-js}\, p^{r_1-js} q^{r_2-j+1}. \end{aligned}$$

Using the relation $\sum_{k=0}^{n} {}^nC_k p^k q^{n-k} = 1$, we obtain

$$P\{g_E \geqslant s\} = \sum_{j\geqslant 1} (-)^{j+1}\, {}^{r-js}C_{j-1} p^{js} q^{j-1} \left(1 + \frac{r-j(s+1)+1}{j} q\right)$$

$$= \sum_{j=1}^{[r/s]} (-)^{j+1} \left(p + \frac{r-js+1}{j} q\right) {}^{r-js}C_{j-1} p^{js} q^{j-1}.$$

This can readily be shown to be identical with the solution using difference equations, as given by Uspensky (1937, p. 77).

I wish to thank Dr F. N. David for suggesting the subject and for advice given in the preparation of this paper.

REFERENCES

DAVID, F. N. (1947). *Biometrika*, **34**, 335.
MARKOFF, A. A. (1913). Supplement to 3rd edition of *Calcul des probabilités*. St Petersbourg.
MOSTELLER, FREDERICK (1941). *Ann. Math. Statist.* **12**, 228.
USPENSKY, J. V. (1937). *Introduction to Mathematical Probability*. McGraw-Hill Book Company.

SUR LES COURBES DE FRÉQUENCE DE K. PEARSON

PAR M. DUMAS, *Ingénieur en chef de l'Artillerie navale* (*Marine nationale française*)

1. K. Pearson a indiqué (1895, 1916) comment il était possible, connaissant les rapports homogènes β_1 et β_2 d'une distribution, de calculer l'équation de la courbe des densités d'une loi de probabilité ajustant cette distribution. Il a dû considérer différents cas, suivant la région du plan des $\beta_1\beta_2$ dans laquelle se situe le point intéressant. Ces régions sont indiquées avec précision dans une figure* (K. Pearson, 1916), reproduite en tête de la Part II des *Tables for Statisticians and Biometricians* (K. Pearson, 1931).

Nous nous proposons dans le présent mémoire d'attirer l'attention sur ce qu'il est utile de compléter les indications de la figure susvisée en traçant une nouvelle courbe. Pour l'exposé correspondant, nous conservons les notations utilisées aux p. lx et suivantes de la Part I, édition de 1930, où se trouve un tableau des équations des courbes des fonctions $y(x)$ de K. Pearson. Nous rappelons que l'équation différentielle servant de point de départ à la théorie est

$$\frac{1}{y}\frac{dy}{dx} = \frac{x-a}{c_0+c_1x+c_2x^2}.$$

2. Celles des lignes tracées sur la figure susvisée qui sont les plus intéressantes à considérer du point de vue des formes des courbes représentatives des fonctions $y(x)$ sont:

(*a*) la droite:
$$2\beta_2-3\beta_1-6 = 0; \tag{1}$$

(*b*) la cubique et la biquadratique ayant respectivement pour équations:

$$\beta_1(\beta_2+3)^2-4(4\beta_2-3\beta_1)(2\beta_2-3\beta_1-6) = 0, \tag{2}$$

$$\beta_1(\beta_2+3)^2(8\beta_2-9\beta_1-12)-4(4\beta_2-3\beta_1)(5\beta_2-6\beta_1-9)^2 = 0. \tag{3}$$

La biquadratique définie par (3) peut être dite *biquadratique de discontinuité des ordonnées*, car on peut établir ce qui suit: si l'on considère les courbes des fonctions $y(x)$ correspondant respectivement aux β_1 et β_2 de deux points voisins de la biquadratique (3) mais situés de part et d'autre de celle-ci, l'une de ces courbes a une ordonnée nulle en un point ayant pour abscisse une solution de l'équation:

$$c_0+c_1x+c_2x^2 = 0, \tag{4}$$

tandis que l'autre courbe a, au point correspondant au précédent, une ordonnée infiniment grande.

C'est ainsi par exemple, que la traversée de la branche marquée VIII sur la figure de la Part II susvisée, fait passer la courbe $y(x)$ du type I_U au type I'_J (comme cela est rappelé sur les figures 2 et 4 ci-jointes), en passant d'ailleurs par le type VIII (Fig. 3).

Il est commode, notamment pour la construction des courbes, de remplacer les équations (2) et (3) par des équations paramétriques, en prenant pour paramètres β_1 et le coefficient c_2 de la quadratique (4). On établit en effet que les équations susvisées sont équivalentes à la relation

$$2(1+5c_2)\beta_2 = 3(1+4c_2)\beta_1+6(1+3c_2), \tag{5}$$

* Cette figure présente notamment l'avantage d'indiquer comme limite de la partie utile du plan des $\beta_1\beta_2$ la droite $\beta_2-\beta_1-1=0$, et d'exclure par suite la droite $4\beta_2-3\beta_1=0$, indiquée dans la figure xxxvi de la Part I des mêmes 'Tables'.

complétée en ce qui concerne la cubique (2) par

$$\beta_1 = -\frac{16c_2(1+3c_2)}{(1+4c_2)^2}; \tag{6}$$

et en ce qui concerne la biquadratique (3) par

$$\beta_1 = \frac{4(1+3c_2)}{(1+4c_2)^2(1+c_2)}. \tag{7}$$

3. L'axe des β_2, la droite $\beta_2-\beta_1-1=0$, la droite (1), et les courbes (2) et (3) limitent différentes régions du plan des $\beta_1\beta_2$; ce sont les seules régions considérées par K. Pearson pour définir ses types de courbes; pour chacune de ces régions, une seule et même expression $y(x)$ est valable. Mais il se trouve que dans certaines de ces régions l'équation correspond à des formes à courbes nettement différentes les unes des autres, car aux points où $y=0$, la pente de la tangente à la courbe $y(x)$ peut être soit *nulle*, soit *infiniment grande*, soit encore, comme état intermédiaire, *finie non nulle*.

Nous avançons que tous les points pour lesquels ces pentes sont finies non nulles, sont ceux situés sur la courbe d'équation

$$\beta_1(\beta_2+3)^2(5\beta_2-6\beta_1-9)-(4\beta_2-3\beta_1)(7\beta_2-9\beta_1-15)^2=0. \tag{8}$$

Soit à démontrer cette proposition.

Rappelons d'abord que la théorie de K. Pearson donne pour les coefficients de la quadratique (4) les expressions

$$c_0=-\mu_2(1+3c_2) \quad \text{et} \quad c_1=-\tfrac{1}{2}\epsilon(1+4c_2)\sqrt{(\mu_2\beta_1)} \tag{9}$$

($\epsilon=\pm1$ ayant le signe de μ_3), auxquelles s'ajoute l'équation (5) donnant c_2 en fonction de β_1 et de β_2.

Puis, appelant x_1 et x_2, avec $x_1<x_2$, les solutions réelles de l'équation (4), c'est à dire

$$\left.\begin{matrix}x_1\\x_2\end{matrix}\right\}=\frac{-c_1\pm\sqrt{(c_1^2-4c_0c_2)}}{2c_2}, \tag{10}$$

posons

$$m_1=\frac{c_1-x_1}{c_2(x_2-x_1)} \quad \text{et} \quad m_2=\frac{x_2-c_1}{c_2(x_2-x_1)}; \tag{11}$$

d'après (9) et (10)

$$\left.\begin{matrix}m_1\\m_2\end{matrix}\right\}=\frac{1}{2c_2}\pm\frac{(1+2c_2)(1+4c_2)\sqrt{\beta_1}}{2c_2\sqrt{[\beta_1(1+4c_2)^2+16c_2(1+3c_2)]}}. \tag{12}$$

Considérons maintenant le cas particulier où la solution $y(x)$ a pour expression

$$y(x)=k(x-x_1)^{m_1}(x-x_2)^{m_2};$$

c'est le cas où, pour $\mu_3>0$, la courbe représentative est du type VI de K. Pearson. Dans ce cas

$$\frac{dy(x)}{dx}=\frac{k}{c_2}(x-c_1)(x-x_1)^{m_1-1}(x-x_2)^{m_2-1}.$$

Manifestement, on ne peut avoir à la fois $y(x)$ nul et $\dfrac{dy(x)}{dx}$ fini non nul que pour $x=x_1$ (ou $x=x_2$), à condition d'avoir en même temps $m_1=1$ (ou $m_2=1$), c'est à dire, d'après (12), à condition d'avoir

$$\left(1-\frac{1}{2c_2}\right)^2=\frac{\beta_1(1+2c_2)^2(1+4c_2)^2}{4c_2^2[\beta_1(1+4c_2)^2+16c_2(1+3c_2)]}.$$

Si dans cette expression on porte la valeur de c_2 déduite de (5), on trouve (8).

Nous ne croyons pas utile d'indiquer ici la discussion complète, seule capable de montrer que la courbe d'équation (8) est valable dans tous les cas et qu'elle est la seule valable.

4. L'équation (8) est l'équation d'une nouvelle biquadratique qui peut être dite, pour la distinguer de celle qui a pour équation (3), *biquadratique de discontinuité des dérivées* en raison de la propriété qui a conduit à la considérer. Ses équations paramètriques sont l'équation (5) complétée par

$$\beta_1 = \frac{2(1+3c_2)(1-2c_2)^2}{(1+4c_2)^2}.$$

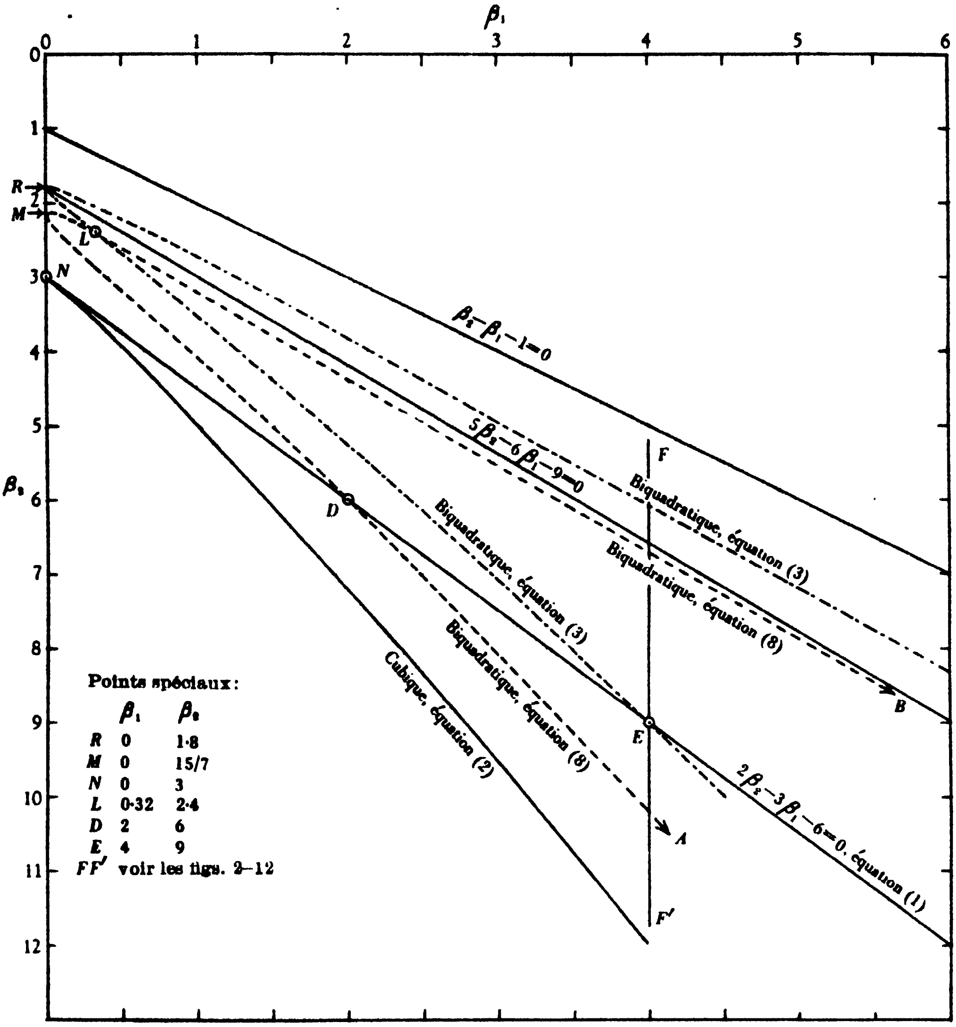


Fig. 1. Limites de régions du plan des $\beta_1\,\beta_2$.

La partie utile de cette biquadratique a deux branches dont l'allure générale est celle des deux branches de la biquadratique (3); l'une d'elles est asymptote à la droite d'équation:

$$5\beta_2 - 6\beta_1 - 9 = 0;$$

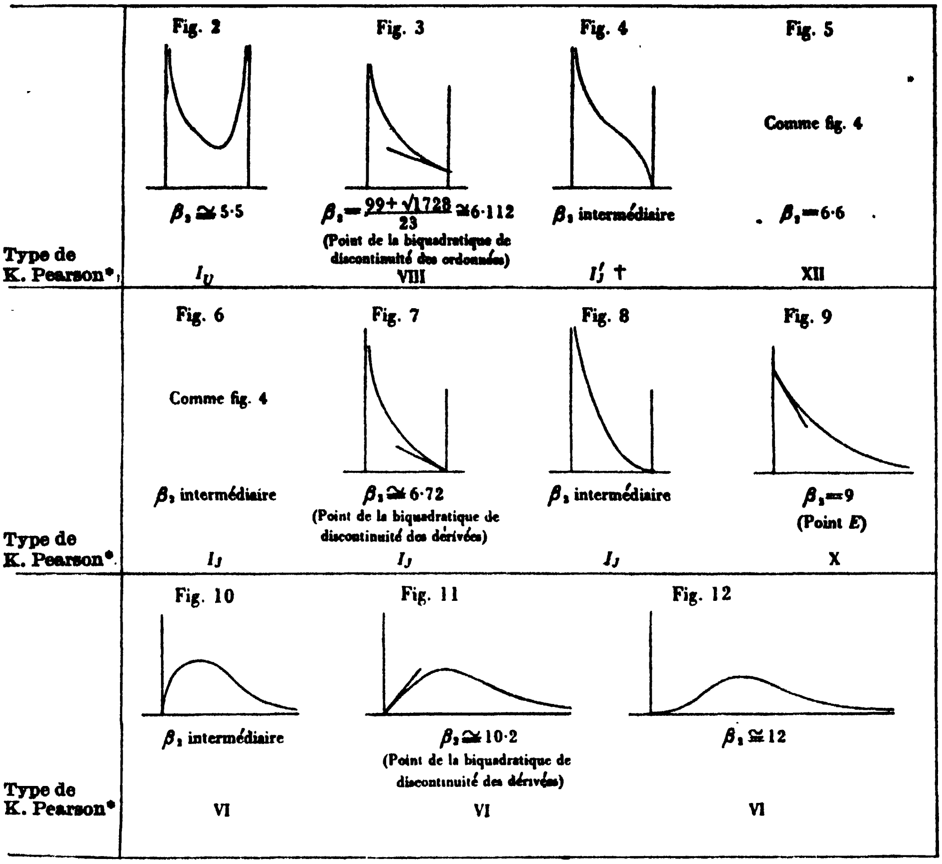
l'autre est pour β_2 infiniment grand, asymptote à

$$\beta_1 = 39{\cdot}2;$$

cette dernière équation est à comparer aux suivantes

$\beta_1 = 32$, qui est celle de l'asymptote à la cubique (2);

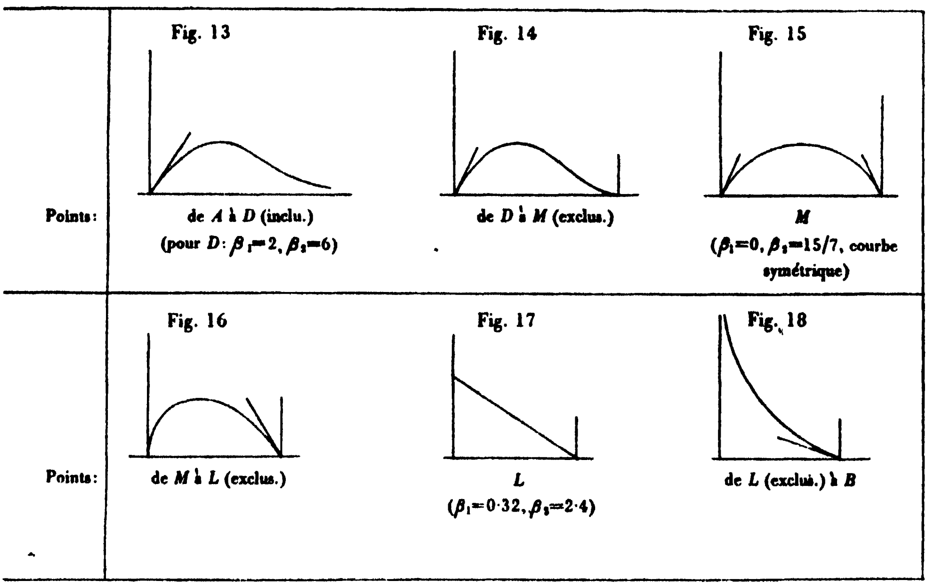
$\beta_1 = 50$, qui est celle d'une asymptote à la biquadratique (3).



Figs. 2–12. Formes des courbes $y(x)$ pour $\beta_1 = 4$ (c'est à dire le long de la section FF' de la figure 1).

* D'après la figure de K. Pearson (1931) visée au § 1 du présent mémoire.

† Nous avons marqué I'_J bien que, sur la figure visée en la note précédente, aucun type ne soit indiqué dans la région correspondante.



Figs. 13–18. Formes des courbes $y(x)$ pour les points de la biquadratique de discontinuité des dérivées.

Comme on le voit sur la figure 1, ces deux branches se raccordent tangentiellement à l'axe des β_2 au point M

$$\beta_1 = 0, \quad \beta_2 = \tfrac{15}{7};$$

l'une d'elles coupe la biquadratique de discontinuité des ordonnées au point

$$\beta_1 = 0{\cdot}32, \quad \beta_2 = 2{\cdot}4,$$

qui est le point L de K. Pearson, tandis que l'autre branche coupe la droite d'équation (1) au point D

$$\beta_1 = 2, \quad \beta_2 = 6.$$

5. Pour donner une idée de l'intérêt de la considération de la biquadratique de discontinuité des dérivées, nous avons établi les figures 2 à 18 ci-jointes. Les figures 2 à 12 indiquent la succession des formes des courbes $y(x)$ que l'on rencontre dans le cas où, β_1 restant constamment égal à 4, β_2 croît depuis les environs de 5·5 jusqu'à ceux de 12; en raison de ces valeurs, le point représentatif reste toujours compris entre le droite limite d'équation

$$\beta_2 - \beta_1 - 1 = 0,$$

et la cubique d'équation (2), et passe par le point

$$\beta_1 = 4, \quad \beta_2 = 9,$$

qui est le point E, point exponentiel, de K. Pearson.

On remarquera qu'au type I_J de K. Pearson (Figs. 6, 7 et 8) correspondent trois courbes d'allures nettement différentes l'une de l'autre, et qu'il en est de même pour le type VI, (Figs. 10, 11 et 12.)

Les figures 13 à 18 indiquent la succession des formes de courbes $y(x)$ que l'on rencontre dans le cas où le point β_1, β_2 suit la biquadratique de discontinuité des dérivées depuis A (Fig. 1) jusqu'à B en passant par M.

Nous comptons publier prochainement un exposé complet de la théorie mathématique conduisant aux fonctions $y(x)$; à cet exposé seront joints un tableau d'environ 34 figures donnant les allures des courbes $y(x)$ pour tous les points possibles du plan des $\beta_1\beta_2$ avec, en particulier, les figures 2 à 18, et toutes précisions analytiques sur les fonctions $y(x)$, notamment sur celles qui correspondent aux points de la biquadratique de discontinuité des dérivées (Figs. 13 à 18).

BIBLIOGRAPHIE

1. PEARSON, K. (1895). Mathematical contributions to the theory of evolution. Skew variation in homogeneous material. *Philos. Trans.* A, **186**, 343–414.
2. PEARSON, K. (1901). *Philos. Trans.* A, **197**, 443–59. Supplément de 1.
3. PEARSON, K. (1916). *Philos. Trans.* A, **216**, 429–57. Second supplément 1.
4. PEARSON, K. (1930). *Tables for Statisticians and Biometricians.* Part I (Edition de 1930).
5. PEARSON, K. (1931). Part II de 4.
6. ELDERTON, W. PALIN (1937). *Frequency-Curves and Correlation.* Cambridge University Press.

THE DISTRIBUTION OF THE EXTREME DEVIATE FROM THE SAMPLE MEAN AND ITS STUDENTIZED FORM*

By K. R. NAIR

1. Introduction

Denote by $x_{(1)} \ldots x_{(n)}$ a random sample of n observations drawn from any statistical universe. Let $x_1, \ldots, x_n$ be the same sample arranged in ascending order of magnitude so that x_r is the rth ranked (or ordered) variate in the sample $\{x_{(i)}\}$.

Various authors have studied the sampling distribution of x_r. It takes a simple form when $r = 1$ or n. Thus, if $f(x)$ is the probability function of the universe from which $x_{(i)}$'s are drawn, the probability that $x_n \leqslant X$ is given by

$$P(X) = \left[\int_{-\infty}^{x} f(x)\,dx\right]^n. \tag{1}$$

If the parent universe is normal with mean μ and standard deviation σ

$$f(x) = \frac{1}{\sigma\sqrt{(2\pi)}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]. \tag{2}$$

The probability that $x_n \leqslant X$ is the same as that of

$$\frac{x_n-\mu}{\sigma} \leqslant \frac{X-\mu}{\sigma} = V \tag{3}$$

and may be written
$$P(X) = P(V) = \left(\frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{V} e^{-\frac{1}{2}x^2}dx\right)^n. \tag{4}$$

Tippett (1925) tabulated values of $P(V)$ for n ranging between 3 and 1000. E. S. Pearson supplemented this by a table of percentage points of V, published as Table XXI *bis* in *Tables for Statisticians and Biometricians*, Part II, to facilitate tests of significance of a single outlying observation x_1 or x_n, when both μ and σ are known.

Very often we do not know either μ or σ or both. When μ alone is unknown, Irwin (1925) suggested a test for a single outlier x_n based on the statistic $(x_n - x_{n-1})/\sigma$. If there are k large outliers, his test would involve $(x_{n-k+1} - x_{n-k})/\sigma$ and, for k small outliers, the statistic $(x_{k+1} - x_k)/\sigma$.

Another statistic commonly used to test a single outlier, when μ is unknown, is one based on the range, viz. $(x_n - x_1)/\sigma$ (see 'Student', 1927).

Intuitively, one feels that a better criterion than $(x_n - x_{n-1})/\sigma$ or $(x_n - x_1)/\sigma$ when there is only *one* outlier x_n will be the extreme deviate $(x_n - \bar{x})/\sigma$ proposed by McKay (1935). The range $(x_n - x_1)$ will certainly be a surer guide than $(x_n - x_{n-1})$ or $(x_n - \bar{x})$ if *both* x_1 and x_n are outliers.

In general, if there are k outliers at the lower end and l outliers at the upper end, a suitable criterion will be that based on the difference of the means of the first k and the last l outliers. When the outliers are all at one end, say the first k, we should write $k + l = n$. The distribution of this general statistic and some of its special cases are considered in § 5.

* Part of a thesis approved for the degree of Ph.D. of the University of London.

All these test criteria should be handled with extreme caution in practice. It may be appropriate here to quote Pearson & Chandrasekar's (1936) warning:

> To base the choice of the test of a statistical hypothesis upon an inspection of the observations is a dangerous practice; a study of the configuration of a sample is almost certain to reveal some feature, or features, which are exceptional if the hypothesis is true.... By choosing the feature most unfavourable to the hypothesis out of a very large number of features examined, it will usually be possible to find some reason for rejecting the hypothesis.

Let us now consider the case where both μ and σ are unknown. Their sample estimates are

$$\bar{x} = \sum_1^n (x_i)/n \quad \text{and} \quad s = \sqrt{\left[\sum_1^n (x_i - \bar{x})^2/(n-1)\right]}. \tag{5}$$

If x_n is to be tested as an outlier, the obvious course is to consider the modified McKay criterion $(x_n - \bar{x})/s$. The sampling distribution of this test criterion is not known, but Thompson (1935) suggested an alternative method based on the sampling distribution of

$$\tau_{(i)} = (x_{(i)} - \bar{x})/s, \tag{6}$$

where (see above) $x_{(i)}$ is a *random* member of the sample. He showed that the distribution of τ can be derived from that of t (with $(n-2)$ degrees of freedom) using the relation

$$\tau = \frac{(n-1)\,t}{\sqrt{[nt^2 + n(n-2)]}}. \tag{7}$$

On a critical examination of Thompson's criterion Pearson & Chandrasekar found that it is effective only when there is essentially *one* outlier and not many as Thompson seemed to believe.

Suppose now that μ is known and σ is unknown. Thompson's criterion could be extended to this situation if we calculate

$$\tau'_{(i)} = (x_{(i)} - \mu)/s', \tag{8}$$

where $s' = \sqrt{\left[\sum_1^n (x_i - \mu)^2/n\right]}$. The distribution of τ' can be obtained from that of t (with $(n-1)$ degrees of freedom) using the relation

$$\tau' = \frac{t\sqrt{n}}{\sqrt{(t^2 + n - 1)}}. \tag{9}$$

We now come to the main theme of this paper which is to consider a method of studentization different from Thompson's. If an estimate s_ν of the unknown σ is available with ν degrees of freedom *independent* of the sample $\{x_{(i)}\}$, a test for a single outlier can be made, using $(x_n - \bar{x})/s_\nu$ if μ is unknown, and using $(x_n - \mu)/s_\nu$ if μ is known. The distributions of these studentized test criteria can be derived by Hartley's (1944) method.

To test whether there are two outliers, one at each end, the studentized range $(x_n - x_1)/s_\nu$ will be a suitable criterion. Pearson & Hartley (1943) have prepared the necessary tables for this test.

In this paper attention is mainly concentrated on the McKay statistic, $u = (x_n - \bar{x})/\sigma$ (or $(\bar{x} - x_1)/\sigma$) and its studentized form $(x_n - \bar{x})/s_\nu$ (or $(\bar{x} - x_1)/s_\nu$). It is shown that the distribution of u can be obtained by a more direct method than was employed by McKay and that it can be reduced to certain integrals which have recently been termed G-functions by Godwin (1945) in his representation of the distribution of the mean deviation. Tables of the probability integral of u and of its studentized form have been prepared.

Apart from serving as a criterion for rejection of an outlying observation in a 'normal' sample when μ and σ are unknown, the studentized form of u has useful application in judging the significance of a single outstanding treatment (best or worst) in a group of treatments tried out in a designed experiment. Some illustrations are given in the paper.

2. Distribution of the extreme deviate

Assuming, without loss of generality, that $\mu = 0$ and $\sigma = 1$ in the normal probability function (2), the joint distribution of the ordered variates $x_1, \ldots, x_n$ is

$$\frac{n!}{[\sqrt{(2\pi)}]^n} \exp\left[-\frac{1}{2}\sum_1^n x_i^2\right] \prod_1^n dx_i. \tag{10}$$

Making an orthogonal transformation of $x_1, \ldots, x_n$ into $y_1, \ldots, y_n$ defined by

$$\left.\begin{aligned} y_i &= \frac{1}{\sqrt{[i(i+1)]}}\{ix_{i+1} - (x_i + \ldots + x_1)\} \quad (i = 1, \ldots, (n-1)) \\ &= \frac{z_i}{\sqrt{[i(i+1)]}} \quad \text{(say)}, \\ y_n &= \frac{1}{\sqrt{n}} \sum_1^n (x_i) = \sqrt{n}\,\bar{x}, \end{aligned}\right\} \tag{11}$$

it follows that

$$\sum_1^n x_i^2 = \sum_1^n y_i^2 = \sum_1^{(n-1)} \frac{z_i^2}{i(i+1)} + n\bar{x}^2, \tag{12}$$

$$\prod_1^n dx_i = \prod_1^n dy_i = \frac{1}{(n-1)!} d\bar{x} \prod_1^{(n-1)} dz_i. \tag{13}$$

The joint distribution of $\bar{x}, z_1, \ldots, z_{(n-1)}$ is therefore

$$\frac{n}{[\sqrt{(2\pi)}]^n} \exp\left[-\frac{n\bar{x}^2}{2} - \frac{1}{2}\sum_1^{(n-1)} \frac{z_i^2}{i(i+1)}\right] d\bar{x} \prod_1^{(n-1)} dz_i. \tag{14}$$

Integrating out for $\bar{x}$ from $-\infty$ to $+\infty$, the joint distribution of $z_1, \ldots, z_{(n-1)}$ is

$$\frac{\sqrt{n}}{[\sqrt{(2\pi)}]^{n-1}} \exp\left[-\frac{1}{2}\sum_1^{(n-1)} \frac{z_i^2}{i(i+1)}\right] \prod_1^{(n-1)} dz_i. \tag{15}$$

It will be noticed that

$$z_{(n-1)} = n(x_n - \bar{x}) = nu. \tag{16}$$

To derive the distribution of $z_{(n-1)}$ or of u, we have to integrate out for $z_1, \ldots, z_{(n-2)}$ in (15). The ranges of the z_i are from 0 to $+\infty$ but they are interlinked by the inequalities

$$0 \leqslant z_1 \leqslant \ldots \leqslant z_{(n-2)} \leqslant z_{(n-1)} = nu. \tag{17}$$

At this stage we bring in the G-functions defined by Godwin as follows:

$$G_0(x) = 1, \quad G_r(x) = \int_0^x \exp\left[-\frac{t^2}{2r(r+1)}\right] G_{r-1}(t)\,dt. \tag{18}$$

Integrating out $z_1, \ldots, z_{n-2}$ from (15) between the limits (17), the distribution of u can be written

$$f_n(u) = \frac{n\sqrt{n}}{[\sqrt{(2\pi)}]^{n-1}} \exp\left[-\frac{nu^2}{2(n-1)}\right] G_{n-2}(nu), \tag{19}$$

and the probability integral of u is

$$P_n(u) = \int_0^u f_n(u)\,du = \frac{\sqrt{n}}{[\sqrt{(2\pi)}]^{n-1}} G_{n-1}(nu). \tag{20}$$

Proceeding by an entirely different method, McKay derived an expression for $P_n(u)$ in the form

$$P_n(u) = \exp\left[-\frac{1}{2n}\frac{d^2}{du^2}\right]\left(\frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^u e^{-\frac{1}{2}x^2}\,dx\right)^n, \tag{21}$$

and obtained a recurrence formula

$$f_n(u) = P_n'(u) = \frac{n}{\sqrt{(2\pi)}} \exp\left[-\frac{nu^2}{2(n-1)}\right] \sqrt{\frac{n}{(n-1)}}\, P_{n-1}\left(\frac{nu}{n-1}\right). \tag{22}$$

The integral-power appearing on the right-hand side of (21) is the probability integral of $(x_n-\mu)/\sigma$ given in (4), revealing an interesting connexion between the probability integrals of $(x_n-\bar{x})/\sigma$ and $(x_n-\mu)/\sigma$.

McKay did not attempt to tabulate values of $P_n(u)$, but for getting the upper 5 or 1 % level of u, that is, when $P_n(u) = 0\cdot95$ or $0\cdot99$, suggested the approximation

$$P_n(u) \sim 1-\frac{n}{\sqrt{(2\pi)}}\int_{u\sqrt{[n/(n-1)]}}^{\infty} e^{-\frac{1}{2}t^2}\,dt. \tag{23}$$

With the help of exact values of $P_n(u)$ tabulated in the next section, it has been possible to examine the closeness of this approximation. The agreement is remarkably good.

He also gave an approximate formula when u is very small, viz.

$$P_n(u) \sim \frac{n}{2}n^{\frac{1}{2}}\left(\frac{u}{\sqrt{(2\pi)}}\right)^{n-1} \quad (n>2). \tag{24}$$

This does not give a good agreement with the exact values.

A better approximation was found to be

$$P_n(u) \sim \frac{\sqrt{n}}{(n-1)!}\left(\frac{nu}{\sqrt{(2\pi)}}\right)^{n-1} \quad (n \geqslant 2), \tag{25}$$

which could further be improved to

$$P_n(u) \sim \frac{\sqrt{n}}{(n-1)!}\left(\frac{nu}{\sqrt{(2\pi)}}\right)^{n-1}\left\{1-\frac{n(n-1)\,u^2}{2(n+1)}\right\} \quad (n \geqslant 2). \tag{26}$$

A comparison of these three approximations with the exact values has been made for a few small values of u and with different values of n in Table 4.

From (25) and (26) we can derive the following approximation for $G_r(x)$ when x is small:

$$G_r(x) \sim \frac{x^r}{r!}, \tag{27}$$

$$G_r(x) \sim \frac{x^r}{r!}\left\{1-\frac{rx^2}{2(r+1)(r+2)}\right\}. \tag{28}$$

3. Tables of the probability integral of the extreme deviate

The starting-point was manuscript tables of the G_r-functions ($r = 2, \ldots, 8$) prepared by Godwin (1945) and Hartley (1945) in the course of their work on the mean deviation. The functions they had actually tabulated were multiples of the G-functions, namely, $G_r^*(x) = b_r\,G_r(x)$. Keeping $n = r+1$, the probability integral of u given in (20) may be written

$$P_{r+1}(u) = c_r\,G_r^*[(r+1)\,u]. \tag{29}$$

Values of b_r and c_r are given below.

r	b_r	c_r
2	$480\,\pi^{-1}\,2^{-\frac{1}{2}}$	25515518×10^{-10}
3	$480^2\pi^{-\frac{3}{2}}\,2^{-1}$	61380797×10^{-13}
4	$480^3\pi^{-2}\,2^{-\frac{3}{2}}$	14297045×10^{-15}
5	$480^4\pi^{-\frac{5}{2}}\,2^{-3}$	65256785×10^{-18}
6	$480^5\pi^{-3}\,2^{-\frac{9}{2}}$	29368910×10^{-20}
7	$480^6\pi^{-\frac{7}{2}}\,2^{-6}$	13081952×10^{-22}
8	$480^7\pi^{-4}\,2^{-\frac{15}{2}}$	57814607×10^{-25}

The task was therefore to convert the G^*-tables from regular arguments in x to regular arguments in u. In order to avoid excessive interpolation for $x = (r+1)u$ in the G_r^*-tables it was decided to adopt a pivotal interval of 0·05 for u, which required selecting for x the pivotal intervals 0·15, 0·20, 0·25, 0·30, 0·35, 0·40 and 0·45 for $r = 2, 3, \ldots, 8$ respectively. The x-interval in Godwin & Hartley's tables was 0·05 for $r = 2, 3, 4$; and 0·10 for $r = 5, 6, 7, 8$. At the intervals 0·35 and 0·45 for x in G_6 and G_8, the half-way point formula of Lagrangian interpolation was applied on the 0·1 interval values of $G_6^*(x)$ and $G_8^*(x)$. Values of $P_n(u)$ for $n = 3$ to 9 were then obtained at intervals of 0·05 and then subtabulated on the National Accounting Machine at intervals of 0·01. These values correct to six decimal places are given in Table 1 at the end of the paper. All the calculations were undertaken by the National Physical Laboratory, Mathematical Division, under the direction of Dr Goodwin and Mr Vickers.

Table 2, printed after Table 1 at the end of the paper, gives percentage points of the extreme deviate, at twelve different levels. These were calculated by interpolation from Table 1.

We are now in a position to examine the adequacy of the approximations (23)–(26). Taking (23) first, the following Table 3 shows the exact and approximate values of the upper 5 and 1 % points.

Table 3. *Comparison of exact and approximate upper percentage points of u*

n	5 %		1 %	
	Exact	Approx. (23)	Exact	Approx. (23)
3	1·7375	1·7376	2·2152	2·2152
4	1·9409	1·9413	2·4310	2·4310
5	2·0801	2·0807	2·5743	2·5743
6	2·1843	2·1854	2·6794	2·6795
7	2·2667	2·2683	2·7613	2·7615
8	2·3344	2·3364	2·8279	2·8281
9	2·3916	2·3940	2·8837	2·8839

The agreement is very good. This high accuracy of his formula was apparently not known to McKay and has now been established by computation of the exact results. It may enable us to extend the range of the present Table 2 for obtaining upper percentage points of u for sample sizes beyond 9.

Approximations (24), (25) and (26) were compared with the exact value of $P_n(u)$ for $u = 0{\cdot}05$, 0·10 and 0·20 and $n = 3, 4, 5$. The results are given in Table 4 where (*a*), (*b*), (*c*) stand for the approximations (24), (25), (26) respectively and (*d*) stands for the exact value. (*a*) and (*b*) are identical when $n = 3$, but it will be seen that when n exceeds 3, (*b*) gives much closer agreement than (*a*).

With the help of the approximations (23) and (26) it may also be possible to extend Godwin & Hartley's tables of upper and lower percentage points of the mean deviation for sample size beyond 10.

To illustrate the use of Table 1 or 2 in testing the significance of an outlying observation an example is given below.

Example. This is taken from McKay's paper. In the course of routine testing of a standard leather product of a tannery, five parallel tests yielded the following values for the hide substance content of the leather specimens:

32·44, 36·45, 39·64, 40·13, 41·09.

The first observation appears unduly low.

Long experience of the product in question has established a value of 2·2 for the standard deviation σ. We get

$$u = (37{\cdot}95 - 32{\cdot}44)/2{\cdot}2 = 2{\cdot}50.$$

Value of P_5 (2·50) found from Table 1 is 0·987 031. The probability of getting a value 2·50 or larger for u is thus about 0·013. If we use Table 2 only, we see that this probability is between 0·025 and 0·01. On the 5 % level of significance, we conclude that the observation 32·44 is anomalous.

Table 4. *Comparison of approximate and exact values of $P_n(u)$ when u is small*

n	Method	$u = 0{\cdot}05$	$u = 0{\cdot}10$	$u = 0{\cdot}20$
3	(a)	0·003 101	0·012 405	0·049 620
	(b)	0·003 101	0·012 405	0·049 620
	(c)	0·003 095	0·012 312	0·048 131
	(d)	0·003 095	0·012 312	0·048 166
4	(a)	0·000 127	0·001 016	0·008 127
	(b)	0·000 169	0·001 355	0·010 836
	(c)	0·000 169	0·001 338	0·010 316
	(d)	0·000 169	0·001 338	0·010 334
5	(a)	0·000 004	0·000 071	0·001 133
	(b)	0·000 009	0·000 148	0·002 360
	(c)	0·000 009	0·000 145	0·002 202
	(d)	0·000 009	0·000 145	0·002 210

(a) McKay's approximation (24).
(b) New approximation (25).
(c) New approximation (26).
(d) Exact value.

4. The studentized integral of the extreme deviate

We have seen that in order to apply the test for the extreme deviate $x_n - \bar{x}$, it is necessary to know the standard deviation σ of the parent normal population. We shall now assume that an estimate s_ν of the unknown σ, equal to the square-root of a variance based on ν degrees of freedom and *independent* of the sample $(x_1, \ldots, x_n)$ is available.

Let ${}_\nu P_n(Q)$ denote the probability that $u_\nu = \dfrac{x_n - \bar{x}}{s_\nu} \leqslant Q$. When $\nu \to \infty$ this tends to the probability $P_n(Q)$ that $u = \dfrac{x_n - \bar{x}}{\sigma} \leqslant Q$ given by (20).

Using Hartley's expansion for studentized integrals up to terms in ν^{-2} we may write

$$ {}_\nu P_n(Q) = a_0 + a_1/\nu + a_2/\nu^2, \tag{30}$$

where
$$a_0 = P_n(Q),$$

and
$$a_1 = \tfrac{1}{4}\{Q^2 P_n''(Q) - Q P_n'(Q)\}, \tag{31}$$

$$a_2 = \tfrac{1}{96}\{3Q^4 P_n^{\mathrm{iv}}(Q) - 2Q^3 P_n'''(Q)\} - \tfrac{1}{8}a_1. \tag{32}$$

Godwin & Hartley had prepared manuscript tables of $G_r^{*\prime}(x)$, the first derivative of $G_r^*(x)$. Values of $P'_{r+1}(Q)$ could be calculated with the help of these tables, using the formula

$$P'_{r+1}(Q) = \frac{24}{W}(r+1)\,c_r\,G_r^{*\prime}(x), \tag{33}$$

where $x \doteq (r+1)\,Q = nQ$ and W is the interval of x in $G_r^{*\prime}(x)$.

Tables of $P'_n(Q)$ for $n = 3, \ldots, 9$ were prepared at the pivotal interval 0·05 for Q just in the same way as $P_n(u)$ was calculated from $G_r^*(x)$.

Values of the 2nd, 3rd and 4th derivatives of $P_n(Q)$ were obtained from $P'_n(Q)$ by numerical differentiation on the National Accounting Machine. Substituting these in (31) and (32), values of a_1 and a_2 were obtained at a pivotal interval of 0·05. Table 5, printed at the end of the paper, gives the values of a_0, a_1 and a_2 at 0·20 interval for the argument Q. Values of ${}_\nu P_n(Q)$ can easily be read off from this table using formula (30). For sake of convenience in making tests of significance, 5 and 1 % points have been calculated in Table 6, which is also reproduced at the end of the paper.

Since numerical differentiation had to be used in calculating even a_1 and a_2, it was not feasible to bring in the next higher terms $a_3\nu^{-3}$ and $a_4\nu^{-4}$ for calculating ${}_\nu P_n(Q)$ as that would have involved calculation of the 5th to 8th derivatives of $P_n(Q)$ by numerical differentiation.

Values of a_1 are given in Table 5 to four decimals for all values of n except $n = 3$, for which five decimals have been supplied. The maximum error is not more than 2 in the last place.

Values of a_2 are given to three decimals for all values of n except $n = 9$, for which only two decimals are retained, when Q exceeds 1·00. The maximum error is not more than 4 in the last place.

This degree of error is inherent in the method of numerical differentiation which was used, but will not seriously affect values of ${}_\nu P_n(Q)$ when ν is moderately large. It is essential, however, to check against systematic errors and for this purpose the following checks proved quite useful:

(i) When $n = 3$, it is comparatively easy to calculate the exact values of a_1 and a_2 directly from the expressions

$$a_1 = \frac{3\sqrt{3}}{8\pi}\,Q\left\{3Qe^{-3Q^2} - (3Q^2+2)\,e^{-\frac{3}{2}Q^2}\int_0^{\frac{3}{2}Q} e^{-t^2}\,dt\right\},$$

$$a_2 = \frac{3\sqrt{3}}{256\pi}\,Q^3\left\{3Q(189Q^2-16)\,e^{-3Q^2} - (27Q^4-42Q^2-8)\,e^{-\frac{3}{2}Q^2}\int_0^{\frac{3}{2}Q} e^{-t^2}\,dt\right\} - \tfrac{1}{8}a_1.$$

In the table below a small panel of values has been calculated from the above exact formulae and compared with the values given in Table 5. The agreement is very satisfactory.

Table 7

Q	a_1		a_2	
	Exact	Approximate	Exact	Approximate
1	−0·38720	−0·38720	+0·275	+0·275
2	−0·25540	−0·25540	−0·550	−0·551
3	−0·01866	−0·01865	−0·324	−0·323
4	−0·00022	−0·00021	−0·014	−0·014

(ii) It is easy to see by partial integration that the expected value of $(x_n - \bar{x})/\sigma$ or $(\bar{x} - x_1)/\sigma$ can be expressed in terms of integrals of a_1 and a_2 over Q from 0 to ∞. Thus,

$$E\left(\frac{x_n - \bar{x}}{\sigma}\right) = \int_0^\infty Q P'_n(Q)\, dQ$$
$$= -\frac{4}{3}\int_0^\infty a_1 dQ$$
$$= -\frac{32}{25}\int_0^\infty a_2 dQ,$$

where $P'_n(Q)$ is given by (33).

Since the expected values of $x_n - \bar{x}$ and $\bar{x} - x_1$ are equal, the expected value of Range/σ or $(x_n - x_1)/\sigma$ is twice that of the extreme deviate. Hence

$$\text{Mean range} = -\frac{8}{3}\int_0^\infty a_1 dQ$$
$$= -\frac{64}{25}\int_0^\infty a_2 dQ.$$

By Gregory's formula for numerical quadrature we can obtain values of the integrals $\int a_1 dQ$, $\int a_2 dQ$ by simple summation. The value of the mean range for $n = 3$ to 9 thus obtained should agree reasonably with the exact values given in Table XXII of *Tables for Statisticians and Biometricians*, Part II. Table 8 below presents the values of the mean range obtained from a_1 and from a_2, alongside the exact values.

Table 8. *Values of mean range* ($\sigma = 1$)

n	Using a_1	Using a_2	Exact value
3	1·69255	1·691	1·69257
4	2·0552	2·053	2·05875
5	2·3236	2·318	2·32593
6	2·5361	2·518	2·53441
7	2·7081	2·687	2·70436
8	2·8504	2·862	2·84720
9	2·9705	2·99	2·97003

Only as many decimals have been retained in columns (2) and (3) of Table 8 as were available for the values of a_1 and a_2 in Table 5. The agreement is satisfactory for our purpose.

Two examples of the use of studentized probability integral of the extreme deviate are given below.

Example 1. This is taken from Snedecor's *Statistical Methods* (1946, 4th ed., p. 266).

A randomized block experiment with four strains of wheat A, B, C, D and five replications gave the following mean yield in pounds per plot.

A	B	C	D
34·4	34·8	33·7	28·4

An interesting feature is the similarity of the first three means. Analysis of variance showed significant differences among the four means, as Table 9 will show.

The variance ratio F for strains against error is $44{\cdot}82/2{\cdot}19 = 20{\cdot}5$ which is much larger than $F_{0\cdot001} = 10{\cdot}8$. As Snedecor says: 'one suspects that the highly significant F is attributable largely to the small yield of D'. But according to him 'no definite probability statements can be made about contrasts suggested by the data'. What he means perhaps is that a test of significance of the difference between 28·4 and the mean of 34·4, 34·8 and 33·7 by the usual t-test

$$t = \frac{34{\cdot}3 - 28{\cdot}4}{\sqrt{[2{\cdot}19(\frac{1}{15}+\frac{1}{5})]}} = 7{\cdot}7,$$

with 12 degrees of freedom is not valid, as 28·4 is the smallest mean. The appropriate criterion in this situation is the studentized extreme deviate which in this case is

$$\frac{\bar{x} - x_1}{s_\nu} = \frac{32{\cdot}8 - 28{\cdot}4}{\sqrt{(\frac{1}{5} \times 2{\cdot}19)}} = 6{\cdot}7,$$

with $n = 4$ and $\nu = 12$. Referring to Table 5 (p. 142 below) we find that $Q = 6{\cdot}7$ is far beyond the limits of the table showing that it is a very exceptional value.

Table 9. *Analysis of variance*

Source of variation	Degrees of freedom	Sum of squares	Variance
Blocks	4	21·46	5·36
Strains	3	134·45	44·82
Error	12	26·26	2·19
Total	19	182·17	—

Having thus concluded that the smallest mean 28·4 is significantly smaller than the other three means, we are justified in saying that D is definitely inferior to A, B and C. Of course, as a general rule, no categorical conclusions can be drawn from the results of a single experiment and only by repeating the experiment a number of times and seeing whether D turns out to give the smallest mean every time can our conclusion from the first results be definitely established.

Example 2. This is artificially built up from Example 1 by changing the error variance from 2·19 to 13·00. The latter gives a standard error per plot of 11 % which is not too excessive a value to expect in ordinary field experiments. The variance ratio for strains against this error is $F = 44{\cdot}82/13{\cdot}00 = 3{\cdot}45$ which is not significant at the 5 % level. The conclusion is therefore that there are no significant differences among the means of the four strains A, B, C, D.

But if we compare the smallest mean against the general mean and calculate the studentized extreme deviate

$$\frac{\bar{x} - x_1}{s_\nu} = \frac{32{\cdot}8 - 28{\cdot}4}{\sqrt{(\frac{1}{5} \times 13{\cdot}00)}} = 2{\cdot}7,$$

we find from Table 6 (see p. 143 below) that the probability of getting this or a larger value when $n = 4, \nu = 12$ lies between 0·05 and 0·01. On the 5 % level, therefore, 28·4 is significantly smaller than the general mean, indicating that D is inferior to A, B and C. Although this is an artificially constructed example, it helps to illustrate a situation which may occasionally arise.

5. Use of G-functions in the distributions of a class of statistics based on ordered variates

Godwin introduced the G-functions primarily to obtain the distribution of the mean deviation from the mean. We have seen its use in the distribution of the extreme deviate from the mean. We shall now consider certain other statistics based on the ordered variates $x_1, \ldots, x_n$ in whose sampling distributions the G-function appears. The distributions derived are in the form of single or double integrals involving the G-functions in the integrand. Only by numerical integration using the G-function tables can the probability integrals of these distributions be calculated. In many cases where only upper and lower percentage points are required, the approximations (27) and (28) to the G-functions may be quite adequate.

Let us consider a general statistic

$$\delta = (x_n + \ldots + x_{n-l+1})/l - (x_1 + \ldots + x_k)/k, \tag{34}$$

where $x_1 \leqslant \ldots \leqslant x_n$ and $k + l \leqslant n$.

When $k = l$, δ is same as the statistic suggested by Jones (1946) for a quick estimation of σ in certain situations where (say) only 5 % of the observations at the two ends are available. As special cases of this statistic we have the range and the mean deviation from the median (see, for example, Nair, 1947).

When $k + l = n$, δ may be written in three different forms:

$$\begin{aligned}\delta &= \frac{1}{(n-k)}(x_n + \ldots + x_{k+1}) - \frac{1}{k}(x_1 + \ldots + x_k)\\ &= \frac{n}{k}\{(x_n + \ldots + x_{k+1})/(n-k) - \bar{x}\}\\ &= \frac{n}{(n-k)}\{\bar{x} - (x_1 + \ldots + x_k)/k\}.\end{aligned} \tag{35}$$

When $k = (n-1)$ or 1, the δ of (35) reduces to the extreme deviate $x_n - \bar{x}$ or $\bar{x} - x_1$, multiplied by the factor $n/(n-1)$.

To obtain the distribution of δ of (34), we start off by introducing the following variates:

$$\left.\begin{aligned}\bar{x} &= \frac{1}{n}(x_1 + \ldots + x_n),\\ \gamma &= \frac{(x_n + \ldots + x_{n-l+1}) + (x_1 + \ldots + x_k)}{(k+l)} - \frac{(x_{k+1} + \ldots + x_{n-l})}{n-k-l},\\ u_i &= \{ix_{i+1} - (x_i + x_{i-1} + \ldots + x_1)\} \quad (i = 1, \ldots, (k-1)),\\ v_i &= \{ix_{i+k+1} - (x_{i+k} + \ldots + x_{k+1})\} \quad (i = 1, \ldots, (n-k-l-1)),\\ w_i &= \{(x_n + \ldots + x_{n-i+1}) - ix_{n-i}\} \quad (i = 1, \ldots, (l-1)).\end{aligned}\right\} \tag{36}$$

Dividing the n new variates, $\delta, \bar{x}, \gamma, u_i, v_i, w_i$ by the square root of the sum of the coefficients of x on the right-hand side, we get the following orthogonal system of transformed variates

$$\begin{aligned}
\delta' &= \sqrt{\frac{kl}{(k+l)}}\,\delta, \\
\bar{x}' &= \sqrt{n}\,\bar{x}, \\
\gamma' &= \sqrt{\frac{(k+l)(n-k-l)}{n}}. \\
u_i' &= \frac{u_i}{\sqrt{[i(i+1)]}} \quad (i = 1, \ldots, (k-1)), \\
v_i' &= \frac{v_i}{\sqrt{[i(i+1)]}} \quad (i = 1, \ldots, (n-k-l-1)), \\
w_i' &= \frac{w_i}{\sqrt{[i(i+1)]}} \quad (i = 1, \ldots, (l-1)).
\end{aligned} \tag{37}$$

It follows that

$$\begin{aligned}
\sum_1^n (x_i^2) &= \bar{x}'^2 + \delta'^2 + \gamma'^2 + \sum_1^{k-1} u_i'^2 + \sum_1^{n-k-l-1} v_i'^2 + \sum_1^{l-1} w_i'^2 \\
&= n\bar{x}^2 + \frac{kl}{(k+l)}\delta^2 + \frac{(k+l)(n-k-l)}{n}\gamma^2 + \sum_1^{k-1} \frac{u_i^2}{i(i+1)} + \sum_1^{n-k-l-1} \frac{v_i^2}{i(i+1)} + \sum_1^{l-1} \frac{w_i^2}{i(i+1)},
\end{aligned} \tag{38}$$

$$\begin{aligned}
\prod_1^n dx_i &= d\bar{x}'\,d\delta'\,d\gamma' \prod_1^{k-1} du_i' \prod_1^{n-k-l-1} dv_i' \prod_1^{l-1} dw_i' \\
&= \frac{d\bar{x}\,d\delta\,d\gamma \prod_1^{k-1} du_i \prod_1^{n-k-l-1} dv_i \prod_1^{l-1} dw_i}{(k-1)!\,(l-1)!\,(n-k-l-1)!}.
\end{aligned} \tag{39}$$

The joint distribution of $\bar{x}$, δ, γ, u_i, v_i, w_i can be obtained from the joint distribution (10) of $x_1, \ldots, x_n$ in the form

$$\begin{aligned}
&\frac{n!(2\pi)^{-\frac{1}{2}n}}{(k-1)!\,(l-1)!\,(n-k-l-1)!} \\
&\times \exp\left[-\frac{1}{2}\left\{n\bar{x}^2 + \frac{kl}{(k+l)}\delta^2 + \frac{(k+l)(n-k-l)}{n}\gamma^2 + \sum_1^{k-1}\frac{u_i^2}{i(i+1)} + \sum_1^{n-k-l-1}\frac{v_i^2}{i(i+1)} + \sum_1^{l-1}\frac{w_i^2}{i(i+1)}\right\}\right] \\
&\times d\bar{x}\,d\delta\,d\gamma \prod_1^{k-1} du_i \prod_1^{n-k-l-1} dv_i \prod_1^{l-1} dw_i.
\end{aligned} \tag{40}$$

Integrating out for $\bar{x}$ between the limits $\pm\infty$, the joint distribution of δ, γ, u_i, v_i, w_i is

$$\begin{aligned}
&\frac{n!\,n^{-\frac{1}{2}}(2\pi)^{-\frac{1}{2}(n-1)}}{(k-1)!\,(l-1)!\,(n-k-l-1)!} \\
&\qquad \times \exp\left[-\frac{1}{2}\left\{\frac{kl}{k+l}\delta^2 + \frac{(k+l)(n-k-l)}{n}\gamma^2 + \sum_1^{k-1}\frac{u_i^2}{i(i+1)} + \sum_1^{n-k-l-1}\frac{v_i^2}{i(i+1)} + \sum_1^{l-1}\frac{w_i^2}{i(i+1)}\right\}\right] \\
&\qquad \times d\delta\,d\gamma \prod_1^{k-1} du_i \prod_1^{n-k-l-1} dv_i \prod_1^{l-1} dw_i.
\end{aligned} \tag{41}$$

To integrate out for γ, u_i, v_i and w_i we have to note that corresponding to the inequalities $x_1 \leqslant \ldots \leqslant x_n$ we now have

$$\left.\begin{aligned}
&0 \leqslant u_1 \leqslant \ldots \leqslant u_{k-1},\\
&0 \leqslant v_1 \leqslant \ldots \leqslant v_{n-k-l-1},\\
&0 \leqslant w_1 \leqslant \ldots \leqslant w_{l-1},\\
&\frac{1}{k}u_{k-1} + \sum_1^{n-k-l-1} \frac{v_i}{i(i+1)} \leqslant \frac{l\delta}{(k+l)} - \gamma,\\
&\frac{1}{l}w_{l-1} + \frac{1}{(n-k-l)}v_{n-k-l-1} \leqslant \frac{k\delta}{(k+l)} + \gamma.
\end{aligned}\right\} \tag{42}$$

It is extremely difficult to carry out the integration for γ, u_i, v_i and w_i subject to the inequalities (42). In the following special cases, however, the last two inequalities in (42) take a simpler form and the integration can be carried forward to a certain stage in terms of G-functions.

Case I. $k+l = n$. Here γ and v_i disappear and the inequalities (42) reduce to

$$\left.\begin{aligned}
&0 \leqslant u_1 \leqslant \ldots \leqslant u_{k-1},\\
&0 \leqslant w_1 \leqslant \ldots \leqslant w_{l-1},\\
&\frac{1}{k}u_{k-1} + \frac{1}{l}w_{l-1} \leqslant \delta.
\end{aligned}\right\} \tag{43}$$

The joint distribution of δ, u_i and w_i is

$$\frac{n!\,n^{-\frac{1}{2}}(2\pi)^{-\frac{1}{2}(n-1)}}{(k-1)!\,(l-1)!} \exp\left[-\frac{1}{2}\left\{\frac{kl}{n}\delta^2 + \sum_1^{k-1} \frac{u_i^2}{i(i+1)} + \sum_1^{l-1} \frac{w_i^2}{i(i+1)}\right\}\right] d\delta \prod_1^{k-1} du_i \prod_1^{l-1} dw_i. \tag{44}$$

Putting $u_{k-1} = \epsilon$ and writing the last inequality in (43) as $w_{l-1} \leqslant l(\delta - \epsilon/k)$ we can integrate out u_i and w_i from (44) and obtain the distribution of δ in the form

$$f(\delta) = \frac{n!\,n^{-\frac{1}{2}}(2\pi)^{-\frac{1}{2}(n-1)}}{(k-1)!\,(l-1)!} \exp\left[-\frac{kl}{2n}\delta^2\right] \int_0^{k\delta} G_{l-1}\{l(\delta - \epsilon/k)\}\, G'_{k-1}(\epsilon)\, d\epsilon, \tag{45}$$

where
$$G'_{k-1}(\epsilon) = \frac{d}{d\epsilon} G_{k-1}(\epsilon) = \exp\left[-\frac{\epsilon^2}{2k(k-1)}\right] G_{k-2}(\epsilon).$$

If we had written $w_{l-1} = \epsilon$ and $u_{k-1} \leqslant k(\delta - \epsilon/l)$, we should get an alternate form which is the same as (45) with k and l interchanged.

Putting $l = 1$, and $\delta = nu/(n-1)$ in (45) we get the distribution of the extreme deviate $x_n - \bar{x}$ given in (19).

Putting $l = 2$, which is appropriate for the case of two outlying observations in the same direction, the distribution of δ takes the form

$$f(\delta) = \frac{2(n-1)(n-2)\sqrt{n}}{(\sqrt{2\pi})^{n-1}} \exp\left[-\frac{(n-2)}{n}\delta^2\right] \int_0^{\delta} G_{n-3}\{(n-2)\,\theta\}\, e^{-(\delta-\theta)^2}\, d\theta. \tag{46}$$

If $k = l = m$ and $n = 2m$, δ becomes twice the mean deviation from the median, whose distribution has been obtained by Godwin (unpublished).

Case II. $k+l = n-1$. Here v_i's disappear and the inequalities (42) reduce to

$$\left.\begin{aligned}
&0 \leqslant u_1 \leqslant \ldots \leqslant u_{k-1} \leqslant k\left(\frac{l\delta}{n-1} - \gamma\right),\\
&0 \leqslant w_1 \leqslant \ldots \leqslant w_{l-1} \leqslant l\left(\frac{k\delta}{n-1} + \gamma\right).
\end{aligned}\right\} \tag{47}$$

The joint distribution of δ, γ, u_i's and w_i's is

$$\frac{n!\,n^{-\frac{1}{2}}(2\pi)^{-\frac{1}{2}(n-1)}}{(k-1)!\,(l-1)!}\exp\left[-\frac{1}{2}\left\{\frac{kl}{(n-1)}\delta^2+\frac{(n-1)}{n}\gamma^2+\sum_1^{k-1}\frac{u_i^2}{i(i+1)}+\sum_1^{l-1}\frac{w_i^2}{i(i+1)}\right\}\right]$$
$$\times\, d\delta\, d\gamma \prod_1^{k-1} du_i \prod_1^{l-1} dw_i. \tag{48}$$

Integrating out γ, u_i and w_i from (48) using the inequalities (47) we get the distribution of δ in the form

$$f(\delta)=\frac{n!\,n^{-\frac{1}{2}}(2\pi)^{-\frac{1}{2}(n-1)}}{(k-1)!\,(l-1)!}\exp\left[-\frac{kl}{2(n-1)}\delta^2\right]$$
$$\times\int_{-k\delta/(n-1)}^{l\delta/(n-1)} G_{k-1}\left\{k\left(\frac{l\delta}{n-1}-\gamma\right)\right\} G_{l-1}\left\{l\left(\frac{k\delta}{n-1}+\gamma\right)\right\}\exp\left[-\frac{(n-1)}{2n}\gamma^2\right]d\gamma. \tag{49}$$

Putting $k=l=m$ and $n=2m+1$, δ becomes n/m times the mean deviation from the median whose distribution has been obtained by Godwin (unpublished). If $m=1$, in particular, δ becomes the range in a sample of 3 and we get

$$f(\delta)=\frac{\sqrt{3}}{\pi}e^{-\frac{1}{4}\delta^2}\int_{-\frac{1}{2}\delta}^{\frac{1}{2}\delta}e^{-\gamma^2/3}\,d\gamma$$
$$=\frac{2\sqrt{3}}{\pi}e^{-\frac{1}{4}\delta^2}\int_0^{\frac{1}{2}\delta}e^{-\gamma^2/3}\,d\gamma, \tag{50}$$

which is the form given by McKay & Pearson (1933).

Case III. $k+l=n-2$. The inequalities become

$$\left.\begin{array}{l} 0\leqslant u_1\leqslant\ldots\leqslant u_{k-1}\leqslant k\left(\dfrac{l\delta}{n-2}-\gamma-\frac{1}{2}v_1\right), \\ 0\leqslant v_1, \\ 0\leqslant w_1\leqslant\ldots\leqslant w_{l-1}\leqslant l\left(\dfrac{k\delta}{n-2}+\gamma-\frac{1}{2}v_1\right). \end{array}\right\} \tag{51}$$

The distribution of δ comes out in the form of a double integral involving product of two G-functions

$$f(\delta)=\frac{n!\,n^{-\frac{1}{2}}(2\pi)^{-\frac{1}{2}(n-1)}}{(k-1)!\,(l-1)!}\exp\left[-\frac{kl}{2(n-2)}\delta^2\right]$$
$$\times\int_0^{\delta}\int_{-\left(\frac{k\delta}{n-2}-\frac{1}{2}v_1\right)}^{\left(\frac{l\delta}{n-2}-\frac{1}{2}v_1\right)} G_{k-1}\left\{k\left(\frac{l\delta}{n-2}-\gamma-\frac{1}{2}v_1\right)\right\} G_{l-1}\left\{l\left(\frac{k\delta}{n-2}+\gamma-\frac{1}{2}v_1\right)\right\}\exp\left[-\left\{\frac{v_1^2}{4}+\frac{(n-2)\gamma^2}{n}\right\}\right]d\gamma\, dv_1. \tag{52}$$

If $n=2m$ and $k=l=m-1$, δ becomes a special case of Jones's statistic. In particular, if $m=2$, this reduces to the range in samples of 4 and its distribution becomes

$$\frac{6\sqrt{2}}{\pi\sqrt{\pi}}e^{-\frac{1}{4}\delta^2}\int_0^{\delta}e^{-\frac{1}{4}v_1^2}\int_0^{\frac{1}{2}(\delta-v_1)}e^{-\frac{1}{2}x^2}\,dx\,dv_1. \tag{53}$$

I should like to acknowledge warmly the help and guidance I have received from Prof. Pearson and Dr Hartley in the course of this investigation.

My thanks are also due to Dr E. T. Goodwin and Mr T. Vickers of the Mathematics Division of the National Physical Laboratory for their expert help in carrying out the calculations required for Tables 1 and 5.

Table 1. *Probability integral of the extreme deviate* $(x_n - \bar{x})/\sigma$ *or* $(\bar{x} - x_1)/\sigma$ *in 'normal' samples of size* n

u \ n	3	4	5	6	7	8	9
	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}
0·00							
0·01	124						
0·02	496	10					
0·03	1 115	36					
0·04	1 982	87	3				
0·05	3 095	169	9				
0·06	4 453	293	19	1			
0·07	6 056	464	35	2			
0·08	7 901	690	59	5	1		
0·09	9 987	979	95	9	1		
0·10	12 312	1 338	145	16	2		
0·11	14 874	1 777	212	25	3		
0·12	17 671	2 301	299	39	5	1	
0·13	20 701	2 917	410	57	8	1	
0·14	23 960	3 631	549	83	13	2	
0·15	27 446	4 451	720	116	19	3	
0·16	31 156	5 382	927	159	28	5	
0·17	35 086	6 430	1 175	214	39	7	1
0·18	39 234	7 601	1 468	283	55	10	2
0·19	43 595	8 901	1 812	368	75	15	3
0·20	48 166	10 334	2 210	472	101	21	5
0·21	52 943	11 906	2 669	597	134	29	7
0·22	57 922	13 621	3 192	747	175	41	10
0·23	63 099	15 483	3 786	924	225	55	13
0·24	68 469	17 495	4 455	1 132	288	73	18
0·25	74 028	19 663	5 204	1 375	363	96	25
0·26	79 772	21 988	6 039	1 656	453	124	34
0·27	85 696	24 474	6 964	1 978	561	159	45
0·28	91 794	27 124	7 985	2 346	689	202	59
0·29	98 063	29 939	9 106	2 765	838	254	77
0·30	104 497	32 922	10 333	3 237	1 013	317	99
0·31	111 091	36 075	11 670	3 768	1 215	392	126
0·32	117 840	39 398	13 121	4 361	1 448	480	159
0·33	124 738	42 893	14 691	5 021	1 715	585	200
0·34	131 781	46 560	16 384	5 754	2 018	707	248
0·35	138 963	50 400	18 205	6 562	2 363	850	306
0·36	146 279	54 412	20 157	7 451	2 752	1 015	375
0·37	153 723	58 597	22 244	8 426	3 188	1 205	456
0·38	161 290	62 954	24 469	9 490	3 677	1 423	551
0·39	168 975	67 481	26 835	10 649	4 221	1 672	662
0·40	176 771	72 179	29 346	11 906	4 825	1 954	791
0·41	184 674	77 045	32 004	13 266	5 493	2 273	940
0·42	192 677	82 079	34 813	14 734	6 228	2 631	1 111
0·43	200 777	87 277	37 773	16 313	7 036	3 033	1 307
0·44	208 966	92 638	40 887	18 007	7 921	3 482	1 530
0·45	217 239	98 159	44 157	19 821	8 886	3 981	1 783
0·46	225 591	103 838	47 584	21 758	9 936	4 535	2 069
0·47	234 017	109 672	51 169	23 822	11 076	5 147	2 390
0·48	242 512	115 658	54 913	26 015	12 309	5 820	2 751
0·49	251 069	121 793	58 816	28 342	13 640	6 560	3 154
0·50	259 684	128 073	62 880	30 805	15 073	7 370	3 602

Table 1 (cont.). *Probability integral of the extreme deviate* $(x_n - \bar{x})/\sigma$ *or* $(\bar{x} - x_1)/\sigma$ *in 'normal' samples of size n*

u \ n	3	4	5	6	7	8	9
	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}
0·50	259 684	128 073	62 880	30 805	15 073	7 370	3 602
0·51	268 351	134 495	67 103	33 407	16 611	8 254	4 099
0·52	277 066	141 054	71 487	36 151	18 259	9 216	4 649
0·53	285 823	147 748	76 029	39 039	20 021	10 260	5 255
0·54	294 616	154 572	80 730	42 073	21 899	11 391	5 922
0·55	303 442	161 522	85 589	45 255	23 899	12 612	6 652
0·56	312 295	168 594	90 604	48 587	26 023	13 928	7 450
0·57	321 170	175 782	95 774	52 069	28 275	15 343	8 321
0·58	330 063	183 084	101 097	55 705	30 657	16 860	9 267
0·59	338 969	190 494	106 572	59 493	33 173	18 484	10 294
0·60	347 883	198 008	112 195	63 436	35 825	20 218	11 404
0·61	356 801	205 621	117 965	67 534	38 616	22 066	12 602
0·62	365 720	213 328	123 879	71 786	41 549	24 032	13 893
0·63	374 633	221 124	129 935	76 193	44 625	26 118	15 279
0·64	383 538	229 006	136 129	80 754	47 846	28 329	16 766
0·65	392 430	236 967	142 459	85 468	51 214	30 668	18 356
0·66	401 305	245 003	148 921	90 334	54 730	33 137	20 054
0·67	410 159	253 110	155 511	95 352	58 396	35 740	21 863
0·68	418 989	261 283	162 227	100 521	62 213	38 478	23 787
0·69	427 791	269 516	169 064	105 838	66 180	41 355	25 830
0·70	436 562	277 805	176 019	111 302	70 300	44 373	27 995
0·71	445 298	286 145	183 088	116 912	74 572	47 534	30 285
0·72	453 996	294 532	190 266	122 666	78 996	50 839	32 703
0·73	462 652	302 960	197 550	128 560	83 571	54 290	35 253
0·74	471 264	311 425	204 935	134 593	88 298	57 889	37 936
0·75	479 829	319 923	212 417	140 761	93 176	61 637	40 756
0·76	488 344	328 448	219 992	147 062	98 203	65 535	43 715
0·77	496 805	336 997	227 655	153 493	103 380	69 583	46 814
0·78	505 211	345 564	235 402	160 050	108 703	73 781	50 057
0·79	513 559	354 147	243 228	166 731	114 172	78 131	53 444
0·80	521 847	362 739	251 129	173 531	119 785	82 632	56 978
0·81	530 072	371 337	259 100	180 447	125 540	87 284	60 660
0·82	538 232	379 938	267 137	187 476	131 434	92 086	64 490
0·83	546 325	388 536	275 235	194 613	137 466	97 038	68 471
0·84	554 349	397 129	283 390	201 855	143 632	102 138	72 601
0·85	562 303	405 711	291 597	209 197	149 930	107 386	76 883
0·86	570 184	414 280	299 851	216 636	156 357	112 780	81 315
0·87	577 991	422 831	308 149	224 167	162 910	118 319	85 899
0·88	585 722	431 362	316 484	231 786	169 585	124 000	90 633
0·89	593 376	439 868	324 854	239 488	176 380	129 822	95 517
0·90	600 951	448 346	333 254	247 270	183 290	135 783	100 550
0·91	608 446	456 793	341 679	255 127	190 313	141 880	105 732
0·92	615 860	465 206	350 125	263 054	197 444	148 111	111 061
0·93	623 192	473 581	358 588	271 047	204 680	154 472	116 536
0·94	630 441	481 917	367 063	279 102	212 016	160 961	122 154
0·95	637 606	490 210	375 547	287 214	219 449	167 575	127 913
0·96	644 686	498 457	384 035	295 379	226 974	174 310	133 811
0·97	651 680	506 656	392 523	303 591	234 587	181 164	139 848
0·98	658 588	514 803	401 008	311 848	242 285	188 132	146 021
0·99	665 408	522 895	409 486	320 143	250 062	195 211	152 329
1·00	672 142	530 930	417 952	328 474	257 914	202 397	158 771

Table 1 (cont.). *Probability integral of the extreme deviate* $(x_n - \bar{x})/\sigma$ *or* $(\bar{x} - x_1)/\sigma$ *in 'normal' samples of size n*

u \ n	3	4	5	6	7	8	9
	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}
1·00	672 142	530 930	417 952	328 474	257 914	202 397	158 771
1·01	678 787	538 909	426 404	336 835	265 838	209 687	165 335
1·02	685 344	546 827	434 838	345 222	273 827	217 076	172 023
1·03	691 812	554 683	443 250	353 632	281 879	224 561	178 832
1·04	698 191	562 474	451 636	362 059	289 989	232 136	185 758
1·05	704 481	570 199	459 994	370 499	298 151	239 799	192 798
1·06	710 682	577 856	468 321	378 949	306 362	247 544	199 947
1·07	716 793	585 443	476 613	387 405	314 618	255 368	207 203
1·08	722 815	592 959	484 867	395 861	322 913	263 266	214 561
1·09	728 748	600 402	493 080	404 316	331 244	271 234	222 018
1·10	734 592	607 770	501 250	412 764	339 606	279 267	229 568
1·11	740 346	615 064	509 374	421 202	347 995	287 360	237 209
1·12	746 011	622 280	517 449	429 626	356 406	295 510	244 935
1·13	751 588	629 418	525 472	438 032	364 835	303 712	252 743
1·14	757 077	636 478	533 442	446 418	373 278	311 961	260 627
1·15	762 477	643 458	541 356	454 780	381 732	320 253	268 585
1·16	767 790	650 356	549 212	463 114	390 191	328 583	276 610
1·17	773 015	657 173	557 007	471 417	398 652	336 948	284 700
1·18	778 154	663 908	564 740	479 686	407 110	345 342	292 848
1·19	783 206	670 559	572 408	487 918	415 563	353 761	301 052
1·20	788 172	677 127	580 010	496 110	424 006	362 202	309 305
1·21	793 053	683 611	587 545	504 259	432 435	370 659	317 604
1·22	797 849	690 010	595 009	512 363	440 847	379 128	325 945
1·23	802 561	696 324	602 403	520 419	449 239	387 606	334 322
1·24	807 190	702 552	609 724	528 423	457 606	396 088	342 731
1·25	811 735	708 695	616 971	536 375	465 946	404 570	351 169
1·26	816 199	714 752	624 143	544 271	474 255	413 049	359 629
1·27	820 581	720 723	631 238	552 109	482 530	421 520	368 109
1·28	824 882	726 608	638 256	559 887	490 767	429 979	376 604
1·29	829 104	732 407	645 195	567 603	498 965	438 424	385 109
1·30	833 246	738 120	652 055	575 255	507 120	446 849	393 621
1·31	837 309	743 746	658 834	582 841	515 229	455 252	402 135
1·32	841 295	749 287	665 531	590 359	523 290	463 629	410 647
1·33	845 205	754 743	672 147	597 808	531 300	471 978	419 153
1·34	849 038	760 112	678 680	605 185	539 256	480 293	427 650
1·35	852 796	765 397	685 131	612 491	547 156	488 573	436 133
1·36	856 479	770 596	691 497	619 722	554 998	496 814	444 599
1·37	860 090	775 711	697 779	626 878	562 780	505 014	453 044
1·38	863 628	780 742	703 977	633 958	570 499	513 168	461 465
1·39	867 094	785 688	710 089	640 960	578 154	521 276	469 858
1·40	870 489	790 552	716 117	647 884	585 742	529 333	478 219
1·41	873 814	795 333	722 059	654 727	593 262	537 337	486 546
1·42	877 071	800 031	727 916	661 490	600 712	545 286	494 835
1·43	880 259	804 647	733 688	668 172	608 090	553 178	503 084
1·44	883 381	809 183	739 374	674 771	615 396	561 010	511 288
1·45	886 436	813 637	744 975	681 288	622 627	568 779	519 446
1·46	889 425	818 012	750 490	687 722	629 782	576 485	527 554
1·47	892 351	822 307	755 920	694 071	636 859	584 124	535 610
1·48	895 213	826 524	761 265	700 336	643 859	591 695	543 611
1·49	898 012	830 663	766 525	706 516	650 779	599 197	551 556
1·50	900 750	834 724	771 701	712 611	657 619	606 627	559 440

Table 1 (cont.). *Probability integral of the extreme deviate $(x_n - \bar{x})/\sigma$ or $(\bar{x} - x_1)/\sigma$ in 'normal' samples by size n*

u \ n	3	4	5	6	7	8	9
	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}
1·50	900 750	834 724	771 701	712 611	657 619	606 627	559 440
1·51	903 428	838 710	776 792	718 621	664 377	613 983	567 262
1·52	906 045	842 619	781 800	724 545	671 053	621 265	575 021
1·53	908 604	846 453	786 724	730 383	677 647	628 471	582 713
1·54	911 105	850 214	791 565	736 136	684 156	635 600	590 337
1·55	913 550	853 901	796 324	741 803	690 582	642 650	597 892
1·56	915 938	857 515	801 000	747 384	696 923	649 620	605 374
1·57	918 271	861 058	805 595	752 879	703 179	656 509	612 783
1·58	920 550	864 531	810 109	758 289	709 349	663 317	620 117
1·59	922 775	867 933	814 543	763 613	715 433	670 042	627 375
1·60	924 949	871 266	818 897	768 852	721 431	676 684	634 555
1·61	927 071	874 531	823 172	774 007	727 343	683 241	641 656
1·62	929 142	877 728	827 368	779 076	733 169	689 714	648 677
1·63	931 164	880 859	831 487	784 062	738 908	696 101	655 616
1·64	933 137	883 924	835 529	788 964	744 560	702 402	662 473
1·65	935 062	886 925	839 494	793 783	750 126	708 617	669 246
1·66	936 940	889 862	843 384	798 518	755 605	714 745	675 935
1·67	938 772	892 737	847 199	803 172	760 999	720 786	682 540
1·68	940 559	895 549	850 941	807 743	766 306	726 740	689 058
1·69	942 301	898 300	854 609	812 233	771 527	732 607	695 491
1·70	944 000	900 991	858 205	816 643	776 663	738 387	701 837
1·71	945 656	903 623	861 729	820 973	781 714	744 078	708 095
1·72	947 270	906 197	865 182	825 223	786 680	749 683	714 266
1·73	948 843	908 713	868 566	829 395	791 561	755 200	720 349
1·74	950 376	911 172	871 881	833 488	796 359	760 630	726 344
1·75	951 870	913 577	875 128	837 505	801 073	765 973	732 251
1·76	953 324	915 926	878 308	841 445	805 704	771 229	738 069
1·77	954 741	918 222	881 421	845 309	810 253	776 398	743 799
1·78	956 121	920 465	884 469	849 099	814 720	781 482	749 440
1·79	957 464	922 656	887 453	852 814	819 106	786 479	754 993
1·80	958 772	924 795	890 373	856 456	823 411	791 391	760 458
1·81	960 045	926 885	893 230	860 025	827 637	796 219	765 834
1·82	961 284	928 925	896 026	863 523	831 783	800 961	771 123
1·83	962 489	930 917	898 760	866 950	835 851	805 620	776 324
1·84	963 662	932 861	901 435	870 308	839 842	810 196	781 438
1·85	964 803	934 759	904 051	873 596	843 756	814 689	786 466
1·86	965 913	936 610	906 608	876 816	847 593	819 099	791 407
1·87	966 992	938 417	909 109	879 969	851 356	823 429	796 261
1·88	968 042	940 180	911 553	883 056	855 044	827 677	801 031
1·89	969 062	941 899	913 942	886 077	858 658	831 846	805 715
1·90	970 054	943 576	916 276	889 034	862 200	835 935	810 315
1·91	971 018	945 211	918 557	891 927	865 670	839 945	814 832
1·92	971 954	946 805	920 785	894 757	869 068	843 878	819 265
1·93	972 864	948 359	922 962	897 526	872 397	847 734	823 616
1·94	973 748	949 874	925 087	900 234	875 657	851 514	827 885
1·95	974 607	951 350	927 163	902 882	878 848	855 219	832 073
1·96	975 441	952 789	929 189	905 471	881 972	858 849	836 181
1·97	976 251	954 190	931 168	908 002	885 029	862 405	840 209
1·98	977 037	955 556	933 099	910 476	888 021	865 889	844 159
1·99	977 800	956 886	934 983	912 894	890 949	869 301	848 031
2·00	978 541	958 181	936 822	915 257	893 813	872 642	851 826

Table 1 (cont.). *Probability integral of the extreme deviate* $(x_n - \bar{x})/\sigma$ *or* $(\bar{x} - x_1)/\sigma$ *in 'normal' samples of size n*

u \ n	3	4	5	6	7	8	9
	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}
2·00	978 541	958 181	936 822	915 257	893 813	872 642	851 826
2·01	979 260	959 442	938 616	917 566	896 614	875 913	855 544
2·02	979 958	960 670	940 366	919 821	899 353	879 116	859 188
2·03	980 635	961 866	942 073	922 023	902 032	882 250	862 757
2·04	981 291	963 030	943 738	924 174	904 650	885 317	866 252
2·05	981 928	964 162	945 362	926 274	907 210	888 318	869 675
2·06	982 545	965 265	946 945	928 325	909 712	891 253	873 027
2·07	983 144	966 337	948 488	930 326	912 157	894 124	876 307
2·08	983 724	967 380	949 992	932 280	914 545	896 932	879 518
2·09	984 286	968 395	951 458	934 187	916 879	899 678	882 661
2·10	984 832	969 382	952 886	936 047	919 158	902 362	885 735
2·11	985 360	970 342	954 278	937 862	921 384	904 986	888 742
2·12	985 872	971 275	955 633	939 632	923 558	907 550	891 684
2·13	986 367	972 182	956 954	941 359	925 680	910 056	894 561
2·14	986 847	973 064	958 240	943 043	927 752	912 504	897 373
2·15	987 312	973 921	959 493	944 685	929 775	914 896	900 123
2·16	987 763	974 754	960 712	946 286	931 748	917 232	902 811
2·17	988 198	975 563	961 899	947 846	933 674	919 514	905 438
2·18	988 620	976 349	963 055	949 367	935 553	921 741	908 005
2·19	989 029	977 113	964 180	950 850	937 386	923 916	910 513
2·20	989 424	977 855	965 274	952 294	939 173	926 040	912 963
2·21	989 806	978 575	966 339	953 701	940 917	928 112	915 355
2·22	990 176	979 275	967 375	955 072	942 617	930 134	917 692
2·23	990 534	979 954	968 383	956 407	944 274	932 107	919 974
2·24	990 880	980 613	969 364	957 708	945 890	934 032	922 201
2·25	991 214	981 253	970 317	958 974	947 465	935 910	924 375
2·26	991 538	981 874	971 244	960 207	949 000	937 742	926 498
2·27	991 851	982 476	972 145	961 407	950 496	939 528	928 568
2·28	992 153	983 061	973 022	962 575	951 953	941 270	930 589
2·29	992 445	983 628	973 873	963 712	953 373	942 968	932 560
2·30	992 727	984 178	974 701	964 819	954 756	944 623	934 483
2·31	992 999	984 711	975 505	965 896	956 102	946 236	936 358
2·32	993 263	985 228	976 287	966 943	957 414	947 808	938 187
2·33	993 517	985 730	977 046	967 962	958 691	949 340	939 970
2·34	993 763	986 216	977 784	968 953	959 934	950 832	941 708
2·35	994 000	986 687	978 500	969 916	961 144	952 286	943 402
2·36	994 229	987 144	979 196	970 853	962 322	953 702	945 053
2·37	994 450	987 587	979 871	971 765	963 468	955 081	946 662
2·38	994 663	988 015	980 527	972 650	964 583	956 424	948 230
2·39	994 869	988 431	981 164	973 511	965 668	957 732	949 758
2·40	995 067	988 833	981 782	974 348	966 723	959 005	951 245
2·41	995 259	989 223	982 381	975 161	967 750	960 243	952 694
2·42	995 443	989 600	982 963	975 951	968 748	961 449	954 105
2·43	995 621	989 966	983 528	976 718	969 719	962 622	955 479
2·44	995 793	990 320	984 075	977 464	970 663	963 764	956 817
2·45	995 959	990 662	984 607	978 188	971 581	964 875	958 119
2·46	996 118	990 993	985 122	978 891	972 473	965 955	959 386
2·47	996 272	991 314	985 622	979 574	973 340	967 006	960 620
2·48	996 420	991 625	986 106	980 237	974 183	968 028	961 820
2·49	996 563	991 925	986 576	980 881	975 001	969 021	962 987
2·50	996 701	992 215	987 031	981 505	975 797	969 987	964 123

Table 1 (cont.). *Probability integral of the extreme deviate* $(x_n - \bar{x})/\sigma$ *or* $(\bar{x} - x_1)/\sigma$ *in 'normal' samples of size n*

u \ n	3	4	5	6	7	8	9
	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}
2·50	996 701	992 215	987 031	981 505	975 797	969 987	964 123
2·51	996 833	992 496	987 473	982 112	976 569	970 926	965 228
2·52	996 961	992 768	987 901	982 700	977 320	971 839	966 303
2·53	997 084	993 031	988 315	983 271	978 049	972 726	967 347
2·54	997 202	993 285	988 717	983 825	978 757	973 588	968 363
2·55	997 316	993 530	989 106	984 362	979 444	974 426	969 351
2·56	997 425	993 768	989 483	984 883	980 111	975 240	970 311
2·57	997 531	993 997	989 848	985 389	980 759	976 030	971 244
2·58	997 632	994 219	990 201	985 879	981 387	976 798	972 151
2·59	997 730	994 433	990 544	986 354	981 998	977 544	973 032
2·60	997 824	994 640	990 875	986 815	982 590	978 268	973 889
2·61	997 914	994 840	991 196	987 261	983 164	978 971	974 720
2·62	998 001	995 033	991 506	987 694	983 721	979 653	975 528
2·63	998 084	995 219	991 807	988 113	984 262	980 316	976 313
2·64	998 165	995 399	992 098	988 520	984 786	980 959	977 074
2·65	998 242	995 573	992 379	988 914	985 294	981 583	977 814
2·66	998 316	995 740	992 651	989 295	985 787	982 188	978 533
2·67	998 387	995 902	992 915	989 665	986 265	982 776	979 230
2·68	998 456	996 058	993 169	990 022	986 728	983 346	979 907
2·69	998 521	996 209	993 416	990 369	987 178	983 898	980 563
2·70	998 585	996 355	993 654	990 705	987 613	984 434	981 201
2·71	998 645	996 495	993 884	991 029	988 035	984 954	981 819
2·72	998 703	996 630	994 107	991 344	988 443	985 458	982 419
2·73	998 759	996 761	994 322	991 648	988 839	985 947	983 001
2·74	998 813	996 886	994 530	991 943	989 223	986 421	983 566
2·75	998 865	997 008	994 731	992 228	989 594	986 880	984 113
2·76	998 914	997 125	994 925	992 504	989 954	987 325	984 644
2·77	998 961	997 237	995 112	992 771	990 303	987 756	985 159
2·78	999 007	997 346	995 293	993 029	990 640	988 174	985 658
2·79	999 050	997 451	995 468	993 279	990 967	988 579	986 141
2·80	999 092	997 551	995 637	993 520	991 283	988 971	986 610
2·81	999 132	997 649	995 800	993 754	991 589	989 351	987 064
2·82	999 171	997 742	995 958	993 979	991 885	989 719	987 505
2·83	999 208	997 832	996 110	994 198	992 172	990 075	987 931
2·84	999 243	997 919	996 257	994 408	992 449	990 420	988 344
2·85	999 277	998 003	996 399	994 612	992 717	990 754	988 745
2·86	999 309	998 083	996 535	994 809	992 977	991 077	989 132
2·87	999 340	998 161	996 667	995 000	993 228	991 390	989 508
2·88	999 370	998 235	996 795	995 184	993 471	991 693	989 871
2·89	999 399	998 307	996 917	995 361	993 705	991 986	990 223
2·90	999 426	998 376	997 036	995 533	993 932	992 269	990 564
2·91	999 452	998 442	997 150	995 699	994 152	992 543	990 894
2·92	999 477	998 506	997 260	995 859	994 364	992 809	991 213
2·93	999 501	998 568	997 366	996 014	994 569	993 065	991 522
2·94	999 524	998 627	997 469	996 163	994 767	993 314	991 821
2·95	999 546	998 683	997 567	996 307	994 959	993 554	992 111
2·96	999 567	998 738	997 663	996 446	995 143	993 786	992 390
2·97	999 587	998 790	997 754	996 580	995 322	994 010	992 661
2·98	999 606	998 841	997 842	996 710	995 495	994 227	992 923
2·99	999 625	998 889	997 928	996 835	995 662	994 437	993 176
3·00	999 642	998 936	998 010	996 955	995 823	994 639	993 421

Table 1 (cont.). *Probability integral of the extreme deviate $(x_n - \bar{x})/\sigma$ or $(\bar{x} - x_1)/\sigma$ in 'normal' samples of size n*

u \ n	3	4	5	6	7	8	9
	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}
3·00	999 642	998 936	998 010	996 955	995 823	994 639	993 421
3·01	999 659	998 981	998 088	997 072	995 978	994 835	993 658
3·02	999 675	999 024	998 164	997 184	996 128	995 024	993 887
3·03	999 690	999 065	998 238	997 292	996 273	995 207	994 108
3·04	999 705	999 105	998 308	997 396	996 413	995 383	994 322
3·05	999 719	999 143	998 376	997 497	996 548	995 554	994 528
3·06	999 732	999 179	998 441	997 594	996 678	995 718	994 728
3·07	999 745	999 215	998 504	997 687	996 804	995 877	994 921
3·08	999 757	999 248	998 565	997 777	996 925	996 031	995 107
3·09	999 769	999 281	998 623	997 864	997 043	996 179	995 287
3·10	999 780	999 312	998 679	997 948	997 155	996 322	995 461
3·11	999 791	999 341	998 733	998 029	997 264	996 460	995 629
3·12	999 801	999 370	998 785	998 106	997 369	996 594	995 792
3·13	999 810	999 397	998 834	998 181	997 471	996 722	995 948
3·14	999 820	999 424	998 882	998 253	997 568	996 846	996 099
3·15	999 829	999 449	998 929	998 322	997 662	996 966	996 245
3·16	999 837	999 473	998 973	998 389	997 753	997 082	996 386
3·17	999 845	999 496	999 015	998 454	997 840	997 193	996 522
3·18	999 853	999 519	999 056	998 515	997 925	997 301	996 653
3·19	999 860	999 540	999 096	998 575	998 006	997 404	996 780
3·20	999 867	999 560	999 134	998 632	998 084	997 504	996 902
3·21	999 873	999 580	999 170	998 688	998 159	997 600	997 020
3·22	999 880	999 599	999 205	998 741	998 232	997 693	997 133
3·23	999 886	999 617	999 238	998 792	998 302	997 783	997 243
3·24	999 891	999 634	999 270	998 841	998 369	997 869	997 349
3·25	999 897	999 650	999 301	998 888	998 434	997 952	997 451
3·26	999 902	999 666	999 331	998 934	998 497	998 032	997 549
3·27	999 907	999 681	999 359	998 978	998 557	998 109	997 644
3·28	999 912	999 696	999 387	999 020	998 615	998 184	997 735
3·29	999 916	999 709	999 413	999 060	998 670	998 255	997 823
3·30	999 920	999 723	999 438	999 099	998 724	998 324	997 908
3·31	999 924	999 735	999 462	999 136	998 775	998 391	997 989
3·32	999 928	999 747	999 486	999 172	998 825	998 455	998 068
3·33	999 932	999 759	999 508	999 207	998 873	998 516	998 144
3·34	999 935	999 770	999 529	999 240	998 919	998 575	998 217
3·35	999 939	999 781	999 550	999 272	998 963	998 633	998 287
3·36	999 942	999 791	999 569	999 302	999 005	998 687	998 355
3·37	999 945	999 801	999 588	999 332	999 046	998 740	998 420
3·38	999 948	999 810	999 606	999 360	999 085	998 791	998 483
3·39	999 951	999 819	999 624	999 387	999 123	998 840	998 544
3·40	999 953	999 827	999 640	999 413	999 159	998 887	998 602
3·41	999 956	999 835	999 656	999 438	999 194	998 932	998 658
3·42	999 958	999 843	999 671	999 462	999 228	998 976	998 712
3·43	999 960	999 850	999 686	999 485	999 260	999 018	998 764
3·44	999 962	999 857	999 700	999 507	999 291	999 058	998 813
3·45	999 964	999 864	999 713	999 528	999 320	999 097	998 862
3·46	999 966	999 871	999 726	999 549	999 349	999 134	998 908
3·47	999 968	999 877	999 738	999 568	999 376	999 169	998 952
3·48	999 970	999 883	999 750	999 587	999 403	999 204	998 995
3·49	999 971	999 888	999 761	999 605	999 428	999 237	999 036
3·50	999 973	999 894	999 772	999 622	999 452	999 269	999 076

Table 1 (cont.). *Probability integral of the extreme deviate $(x_n - \bar{x})/\sigma$ or $(\bar{x} - x_1)/\sigma$ in 'normal' samples of size n*

u \ n	3	4	5	6	7	8	9
	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}
3·50	999 973	999 894	999 772	999 622	999 452	999 269	999 076
3·51	999 974	999 899	999 783	999 638	999 476	999 299	999 114
3·52	999 976	999 904	999 792	999 654	999 498	999 328	999 150
3·53	999 977	999 908	999 802	999 669	999 519	999 357	999 185
3·54	999 978	999 913	999 811	999 684	999 540	999 384	999 219
3·55	999 979	999 917	999 820	999 698	999 560	999 410	999 251
3·56	999 981	999 921	999 828	999 711	999 579	999 435	999 282
3·57	999 982	999 925	999 836	999 724	999 597	999 458	999 312
3·58	999 983	999 929	999 843	999 736	999 614	999 481	999 341
3·59	999 984	999 932	999 851	999 748	999 631	999 503	999 369
3·60	999 985	999 935	999 858	999 759	999 647	999 525	999 395
3·61	999 985	999 939	999 864	999 770	999 662	999 545	999 421
3·62	999 986	999 942	999 870	999 780	999 677	999 564	999 445
3·63	999 987	999 945	999 876	999 790	999 691	999 583	999 469
3·64	999 988	999 947	999 882	999 800	999 705	999 601	999 491
3·65	999 988	999 950	999 888	999 809	999 718	999 618	999 513
3·66	999 989	999 952	999 893	999 817	999 730	999 635	999 533
3·67	999 990	999 955	999 898	999 826	999 742	999 651	999 553
3·68	999 990	999 957	999 903	999 834	999 754	999 666	999 573
3·69	999 991	999 959	999 908	999 841	999 765	999 680	999 591
3·70	999 991	999 961	999 912	999 848	999 775	999 694	999 609
3·71	999 992	999 963	999 916	999 855	999 785	999 708	999 625
3·72	999 992	999 965	999 920	999 862	999 795	999 720	999 642
3·73	999 993	999 967	999 924	999 868	999 804	999 733	999 657
3·74	999 993	999 968	999 928	999 874	999 813	999 745	999 672
3·75	999 994	999 970	999 931	999 880	999 821	999 756	999 686
3·76	999 994	999 972	999 934	999 886	999 829	999 767	999 700
3·77	999 994	999 973	999 938	999 891	999 837	999 777	999 713
3·78	999 995	999 974	999 941	999 896	999 844	999 787	999 726
3·79	999 995	999 976	999 943	999 901	999 852	999 796	999 738
3·80	999 995	999 977	999 946	999 906	999 858	999 806	999 749
3·81	999 995	999 978	999 949	999 910	999 865	999 814	999 760
3·82	999 996	999 979	999 951	999 914	999 871	999 823	999 771
3·83	999 996	999 980	999 954	999 918	999 877	999 831	999 781
3·84	999 996	999 981	999 956	999 922	999 883	999 838	999 791
3·85	999 996	999 982	999 958	999 926	999 888	999 846	999 800
3·86	999 997	999 983	999 960	999 929	999 893	999 852	999 809
3·87	999 997	999 984	999 962	999 933	999 898	999 859	999 817
3·88	999 997	999 985	999 964	999 936	999 903	999 866	999 826
3·89	999 997	999 986	999 966	999 939	999 907	999 872	999 834
3·90	999 997	999 987	999 968	999 942	999 912	999 878	999 841
3·91	999 997	999 987	999 969	999 945	999 916	999 883	999 848
3·92	999 998	999 988	999 971	999 947	999 920	999 889	999 855
3·93	999 998	999 989	999 972	999 950	999 924	999 894	999 862
3·94	999 998	999 989	999 974	999 952	999 927	999 899	999 868
3·95	999 998	999 990	999 975	999 955	999 931	999 903	999 874
3·96	999 998	999 990	999 976	999 957	999 934	999 908	999 880
3·97	999 998	999 991	999 977	999 959	999 937	999 912	999 885
3·98	999 998	999 991	999 979	999 961	999 940	999 916	999 890
3·99	999 998	999 992	999 980	999 963	999 943	999 920	999 895
4·00	999 999	999 992	999 981	999 965	999 946	999 924	999 900

Table 1 (cont.). *Probability integral of the extreme deviate $(x_n - \bar{x})/\sigma$ or $(\bar{x} - x_1)/\sigma$ in 'normal' samples of size n*

u \ n	3	4	5	6	7	8	9
	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}
4·00	999 999	999 992	999 981	999 965	999 946	999 924	999 900
4·01	999 999	999 993	999 982	999 966	999 948	999 927	999 905
4·02	999 999	999 993	999 983	999 968	999 951	999 931	999 909
4·03	999 999	999 993	999 983	999 970	999 953	999 934	999 913
4·04	999 999	999 994	999 984	999 971	999 955	999 937	999 917
4·05	999 999	999 994	999 985	999 973	999 958	999 940	999 921
4·06	999 999	999 994	999 986	999 974	999 960	999 943	999 925
4·07	999 999	999 995	999 987	999 975	999 962	999 946	999 928
4·08	999 999	999 995	999 987	999 976	999 964	999 948	999 932
4·09	999 999	999 995	999 988	999 978	999 965	999 951	999 935
4·10	999 999	999 996	999 989	999 979	999 967	999 953	999 938
4·11	999 999	999 996	999 989	999 980	999 969	999 955	999 941
4·12	999 999	999 996	999 990	999 981	999 970	999 957	999 944
4·13	999 999	999 996	999 990	999 982	999 972	999 959	999 946
4·14	999 999	999 996	999 991	999 983	999 973	999 961	999 949
4·15	1·000 000	999 997	999 991	999 984	999 975	999 963	999 951
4·16		999 997	999 992	999 984	999 976	999 965	999 954
4·17		999 997	999 992	999 985	999 977	999 967	999 956
4·18		999 997	999 993	999 986	999 978	999 968	999 958
4·19		999 997	999 993	999 987	999 979	999 970	999 960
4·20		999 997	999 993	999 987	999 980	999 971	999 962
4·21		999 998	999 994	999 988	999 981	999 973	999 964
4·22		999 998	999 994	999 989	999 982	999 974	999 965
4·23		999 998	999 994	999 989	999 983	999 975	999 967
4·24		999 998	999 995	999 990	999 984	999 977	999 969
4·25		999 998	999 995	999 990	999 985	999 978	999 970
4·26		999 998	999 995	999 991	999 986	999 979	999 971
4·27		999 998	999 996	999 991	999 986	999 980	999 973
4·28		999 998	999 996	999 992	999 987	999 981	999 974
4·29		999 998	999 996	999 992	999 988	999 982	999 975
4·30		999 999	999 996	999 993	999 988	999 983	999 977
4·31		999 999	999 996	999 993	999 989	999 983	999 978
4·32		999 999	999 997	999 993	999 990	999 984	999 979
4·33		999 999	999 997	999 994	999 990	999 985	999 980
4·34		999 999	999 997	999 994	999 991	999 986	999 981
4·35		999 999	999 997	999 994	999 991	999 987	999 982
4·36		999 999	999 997	999 995	999 992	999 987	999 983
4·37		999 999	999 997	999 995	999 992	999 988	999 983
4·38		999 999	999 998	999 995	999 992	999 988	999 984
4·39		999 999	999 998	999 995	999 993	999 989	999 985
4·40		999 999	999 998	999 996	999 993	999 990	999 986
4·41		999 999	999 998	999 996	999 994	999 990	999 986
4·42		999 999	999 998	999 996	999 994	999 991	999 987
4·43		999 999	999 998	999 996	999 994	999 991	999 988
4·44		999 999	999 998	999 996	999 995	999 992	999 988
4·45		999 999	999 998	999 997	999 995	999 992	999 989
4·46		999 999	999 998	999 997	999 995	999 992	999 989
4·47		999 999	999 999	999 997	999 996	999 993	999 990
4·48		999 999	999 999	999 997	999 996	999 993	999 991
4·49		999 999	999 999	999 997	999 996	999 993	999 991
4·50		999 999	999 999	999 998	999 996	999 994	999 992

Table 1 (cont.). *Probability integral of the extreme deviate $(x_n - \bar{x})/\sigma$ or $(\bar{x} - x_1)/\sigma$ in 'normal' samples of size n*

u \ n	3	4	5	6	7	8	9
	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}
4·50		999 999	999 999	999 998	999 996	999 994	999 992
4·51		999 999	999 999	999 998	999 997	999 994	999 992
4·52		999 999	999 999	999 998	999 997	999 994	999 992
4·53		999 999	999 999	999 998	999 997	999 995	999 993
4·54		999 999	999 999	999 998	999 997	999 995	999 993
4·55		1·000 000	999 999	999 998	999 997	999 995	999 993
4·56			999 999	999 998	999 997	999 996	999 994
4·57			999 999	999 998	999 998	999 996	999 994
4·58			999 999	999 998	999 998	999 996	999 994
4·59			999 999	999 998	999 998	999 996	999 995
4·60			1·000 000	999 999	999 998	999 996	999 995
4·61				999 999	999 998	999 996	999 995
4·62				999 999	999 998	999 997	999 995
4·63				999 999	999 998	999 997	999 995
4·64				999 999	999 998	999 997	999 996
4·65				999 999	999 999	999 997	999 996
4·66				999 999	999 999	999 997	999 996
4·67				999 999	999 999	999 997	999 996
4·68				999 999	999 999	999 998	999 997
4·69				999 999	999 999	999 998	999 997
4·70				999 999	999 999	999 998	999 997

Table 2. *Percentage points of the extreme deviate $(x_n - \bar{x})/\sigma$ or $(\bar{x} - x_1)/\sigma$*

Size of sample n	Lower percentage points						Upper percentage points					
	0·1	0·5	1·0	2·5	5·0	10·0	10·0	5·0	2·5	1·0	0·5	0·1
3	0·03	0·06	0·09	0·14	0·20	0·29	1·50	1·74	1·95	2·22	2·40	2·78
4	0·09	0·16	0·20	0·27	0·35	0·45	1·70	1·94	2·16	2·43	2·62	3·01
5	0·16	0·25	0·30	0·38	0·47	0·58	1·83	2·08	2·30	2·57	2·76	3·17
6	0·23	0·33	0·38	0·48	0·56	0·68	1·94	2·18	2·41	2·68	2·87	3·28
7	0·30	0·40	0·46	0·56	0·65	0·76	2·02	2·27	2·49	2·76	2·95	3·36
8	0·36	0·47	0·53	0·62	0·72	0·84	2·09	2·33	2·56	2·83	3·02	3·43
9	0·41	0·53	0·59	0·69	0·78	0·90	2·15	2·39	2·61	2·88	3·07	3·48

Table 5. *Auxiliary quantities required for calculating the probability integral of the studentized extreme deviate* $(x_n-\bar{x})/s_\nu$ *or* $(\bar{x}-x_1)/s_\nu$

n \ Q	3			4			5			6			n / Q
	a_0	a_1	a_2	a_0	a_1	a_2	a_0	a_1	a_2	a_0	a_1	a_2	
0·00	0·000000	—	—	0·000000	—	—	0·000000	—	—	0·000000	—	—	0·00
0·20	0·048166	−0·00277	—	0·010334	+0·0063	−0·003	0·002210	+0·0039	−0·001	0·000472	+0·0016	—	0·20
0·40	0·176771	−0·03594	+0·012	0·072179	+0·0196	−0·022	0·029346	+0·0331	−0·028	0·011906	+0·0281	−0·016	0·40
0·60	0·347883	−0·13016	+0·076	0·198008	−0·0243	+0·004	0·112195	+0·0466	−0·076	0·063436	+0·0787	−0·113	0·60
0·80	0·521847	−0·26547	+0·195	0·362739	−0·1622	+0·169	0·251129	−0·0463	+0·038	0·173531	+0·0486	−0·120	0·80
1·00	0·672142	−0·38720	+0·275	0·530930	−0·3459	+0·402	0·417952	−0·2483	+0·381	0·328474	−0·1323	+0·231	1·00
1·20	0·788172	−0·45487	+0·235	0·677127	−0·4989	+0·524	0·580010	−0·4716	+0·721	0·496110	−0·3993	+0·783	1·20
1·40	0·870489	−0·45998	+0·083	0·790552	−0·5748	+0·430	0·716117	−0·6273	+0·792	0·647884	−0·6314	+1·103	1·40
1·60	0·924949	−0·41502	−0·131	0·871266	−0·5686	+0·148	0·818897	−0·6788	+0·530	0·768852	−0·7510	+0·953	1·60
1·80	0·958772	−0·34014	−0·359	0·924795	−0·5018	−0·215	0·890373	−0·6376	+0·059	0·856456	−0·7492	+0·426	1·80
2·00	0·978541	−0·25540	−0·551	0·958181	−0·4036	−0·549	0·936822	−0·5384	−0·441	0·915257	−0·6599	−0·237	2·00
2·20	0·989424	−0·17657	−0·659	0·977855	−0·2994	−0·787	0·965274	−0·4167	−0·835	0·952294	−0·5278	−0·810	2·20
2·40	0·995067	−0·11277	−0·666	0·988833	−0·2062	−0·891	0·981782	−0·2990	−1·058	0·974348	−0·3897	−1·176	2·40
2·60	0·997824	−0·06667	−0·587	0·994640	−0·1323	−0·869	0·990875	−0·2002	−1·107	0·986815	−0·2682	−1·314	2·60
2·80	0·999092	−0·03656	−0·460	0·997551	−0·0794	−0·752	0·995637	−0·1255	−1·014	0·993520	−0·1729	−1·255	2·80
3·00	0·999642	−0·01865	−0·323	0·998936	−0·0445	−0·587	0·998010	−0·0739	−0·836	0·996955	−0·1048	−1·072	3·00
3·20	0·999867	−0·00884	−0·207	0·999560	−0·0234	−0·417	0·999134	−0·0409	−0·625	0·998632	−0·0598	−0·833	3·20
3·40	0·999953	−0·00391	−0·119	0·999827	−0·0115	−0·271	0·999640	−0·0213	−0·432	0·999413	−0·0322	−0·594	3·40
3·60	0·999985	−0·00163	−0·063	0·999935	−0·0054	−0·161	0·999858	−0·0105	−0·276	0·999759	−0·0164	−0·392	3·60
3·80	0·999995	−0·00064	−0·031	0·999977	−0·0024	−0·091	0·999946	−0·0049	−0·162	0·999906	−0·0079	−0·243	3·80
4·00	0·999999	−0·00021	−0·014	0·999992	−0·0010	−0·045	0·999981	−0·0021	−0·090	0·999965	−0·0036	−0·136	4·00

Table 5 (cont.). *Auxiliary quantities required for calculating the probability integral of the studentized extreme deviate* $(x_n - \bar{x})/s_\nu$ *or* $(\bar{x} - x_1)/s_\nu$

n / Q	7			8			9			n / Q
	a_0	a_1	a_2	a_0	a_1	a_2	a_0	a_1	a_2	
0·00	0·000000	—	—	0·000000	—	—	0·000000	—	—	**0·00**
0·20	0·000101	+0·0005	+0·001	0·000021	+0·0002	—	0·000005	—	—	**0·20**
0·40	0·004825	+0·0192	−0·003	0·001954	+0·0117	+0·005	0·000791	+0·0066	+0·008	**0·40**
0·60	0·035825	+0·0842	−0·105	0·020218	+0·0758	−0·073	0·011404	+0·0619	−0·033	**0·60**
0·80	0·119785	+0·1133	−0·246	0·082632	+0·1498	−0·314	0·056978	+0·1642	−0·324	**0·80**
1·00	0·257914	−0·0199	+0·007	0·202397	+0·0775	−0·235	0·158771	+0·1552	−0·457	**1·00**
1·20	0·424006	−0·3018	+0·706	0·362202	−0·1934	+0·516	0·309305	−0·0838	+0·24	**1·20**
1·40	0·585742	−0·5988	+1·320	0·529333	−0·5395	+1·425	0·478219	−0·4615	+1·41	**1·40**
1·60	0·721431	−0·7907	+1·374	0·676684	−0·8028	+1·757	0·634555	−0·7919	+2·08	**1·60**
1·80	0·823411	−0·8384	+0·852	0·791391	−0·9071	+1·313	0·760458	−0·9573	+1·79	**1·80**
2·00	0·893813	−0·7685	+0·051	0·872642	−0·8648	+0·402	0·851826	−0·9493	+0·81	**2·00**
2·20	0·939173	−0·6327	−0·714	0·926040	−0·7312	−0·556	0·912963	−0·8234	−0·34	**2·20**
2·40	0·966723	−0·4780	−1·238	0·959005	−0·5635	−1·266	0·951245	−0·6462	−1·24	**2·40**
2·60	0·982590	−0·3356	−1·488	0·978268	−0·4022	−1·638	0·973889	−0·4678	−1·76	**2·60**
2·80	0·991283	−0·2206	−1·478	0·988971	−0·2684	−1·685	0·986610	−0·3160	−1·88	**2·80**
3·00	0·995823	−0·1364	−1·299	0·994639	−0·1684	−1·518	0·993421	−0·2006	−1·73	**3·00**
3·20	0·998084	−0·0795	−1·035	0·997504	−0·0997	−1·233	0·996902	−0·1201	−1·43	**3·20**
3·40	0·999159	−0·0437	−0·756	0·998887	−0·0558	−0·918	0·998602	−0·0679	−1·08	**3·40**
3·60	0·999647	−0·0229	−0·511	0·999525	−0·0295	−0·633	0·999395	−0·0365	−0·75	**3·60**
3·80	0·999858	−0·0112	−0·324	0·999806	−0·0147	−0·404	0·999749	−0·0185	−0·49	**3·80**
4·00	0·999946	−0·0052	−0·186	0·999924	−0·0070	−0·243	0·999900	−0·0090	−0·30	**4·00**
4·20	0·999980	−0·0024	−0·106	0·999971	−0·0032	−0·134	0·999962	−0·0042	−0·16	**4·20**
4·40	0·999993	−0·0010	−0·054	0·999990	−0·0014	−0·070	0·999986	−0·0018	−0·08	**4·40**
4·60	0·999998	−0·0004	−0·021	0·999996	−0·0006	−0·035	0·999995	−0·0007	−0·04	**4·60**

Table 6 A. *Lower per cent points of the studentized extreme deviate* $(x_n - \bar{x})/s_\nu$ *or* $(\bar{x} - x_1)/s_\nu$

ν \ n	5 %							1 %						
	3	4	5	6	7	8	9	3	4	5	6	7	8	9
10	0·20	0·35	0·46	0·55	0·62	0·69	0·74	0·09	0·19	0·29	0·37	0·43	0·49	0·54
15	0·20	0·35	0·46	0·55	0·63	0·70	0·75	0·09	0·19	0·29	0·37	0·44	0·50	0·56
30	0·20	0·35	0·46	0·56	0·64	0·70	0·77	0·09	0·20	0·29	0·38	0·45	0·51	0·57
∞	0·20	0·35	0·47	0·56	0·65	0·72	0·78	0·09	0·20	0·30	0·38	0·46	0·53	0·59

B. *Upper per cent points of the studentized extreme deviate* $(x_n - \bar{x})/s_\nu$ *or* $(\bar{x} - x_1)/s_\nu$

ν \ n	5 %							1 %						
	3	4	5	6	7	8	9	3	4	5	6	7	8	9
10	2·02	2·29	2·49	2·63	2·75	2·85	2·93	2·76	3·05	3·25	3·39	3·50	3·59	3·67
11	1·99	2·26	2·44	2·58	2·70	2·79	2·87	2·71	3·00	3·19	3·33	3·44	3·53	3·61
12	1·97	2·22	2·40	2·54	2·65	2·75	2·83	2·67	2·95	3·14	3·28	3·39	3·48	3·55
13	1·95	2·20	2·38	2·51	2·62	2·71	2·79	2·63	2·91	3·10	3·24	3·34	3·43	3·51
14	1·93	2·18	2·35	2·48	2·59	2·68	2·76	2·60	2·87	3·06	3·20	3·30	3·39	3·47
15	1·92	2·16	2·33	2·46	2·56	2·65	2·73	2·57	2·84	3·02	3·16	3·27	3·35	3·43
16	1·90	2·14	2·31	2·44	2·54	2·63	2·70	2·55	2·81	3·00	3·13	3·24	3·32	3·39
17	1·89	2·13	2·30	2·42	2·52	2·61	2·68	2·52	2·79	2·97	3·10	3·21	3·29	3·36
18	1·88	2·12	2·28	2·41	2·51	2·59	2·66	2·50	2·77	2·95	3·08	3·18	3·27	3·34
19	1·87	2·11	2·27	2·39	2·49	2·58	2·65	2·49	2·75	2·92	3·06	3·16	3·24	3·31
20	1·87	2·10	2·26	2·38	2·48	2·56	2·63	2·47	2·73	2·91	3·04	3·14	3·22	3·29
24	1·84	2·07	2·23	2·35	2·44	2·52	2·59	2·43	2·68	2·85	2·97	3·07	3·15	3·22
30	1·82	2·04	2·20	2·31	2·40	2·48	2·55	2·38	2·62	2·79	2·91	3·01	3·08	3·15
40	1·80	2·02	2·17	2·28	2·37	2·44	2·51	2·34	2·57	2·73	2·85	2·94	3·02	3·08
60	1·78	1·99	2·14	2·25	2·33	2·41	2·47	2·30	2·52	2·68	2·79	2·88	2·95	3·01
120	1·76	1·97	2·11	2·21	2·30	2·37	2·43	2·25	2·48	2·62	2·73	2·82	2·89	2·95
∞	1·74	1·94	2·08	2·18	2·27	2·33	2·39	2·22	2·43	2·57	2·68	2·76	2·83	2·88

REFERENCES

Godwin, H. J. (1945). On the distribution of the estimate of mean deviation obtained from samples from a normal population. *Biometrika*, **33**, 254.

Hartley, H. O. (1944). Studentization, or the elimination of the standard deviation of the parent population from the random sample distribution of statistics. *Biometrika*, **33**, 173.

Hartley, H. O. (1945). Note on the calculation of the distribution of the estimate of mean deviation in normal samples. *Biometrika*, **33**, 257.

Irwin, J. O. (1925). On a criterion for the rejection of outlying observations. *Biometrika*, **17**, 238.

Jones, A. E. (1946). A useful method for the routine estimation of dispersion from large samples. *Biometrika*, **33**, 274.

McKay, A. T. (1935). The distribution of the difference between the extreme observation and the sample mean in samples of n from a normal universe. *Biometrika*, **27**, 466.

McKay, A. T. & Pearson, E. S. (1933). A note on the distribution of range in samples of n. *Biometrika*, **25**, 415.

Nair, K. R. (1947). A note on the mean deviation from the median. *Biometrika*, **34**, 360.

Pearson, E. S. & Chandrasekar, C. (1936). The efficiency of statistical tools and a criterion for the rejection of outlying observations. *Biometrika*, **28**, 308.

Pearson, E. S. & Hartley, H. O. (1943). Tables of the probability integral of the studentized range. *Biometrika*, **33**, 89.

'Student' (1927). Errors of routine analysis. *Biometrika*, **19**, 151.

Thompson, W. R. (1935). On a criterion for the rejection of observations and the distribution of the ratio of deviation to sampling standard deviation. *Ann. Math. Statist.* **6**, 214.

Tippett, L. H. C. (1925). On the extreme individuals and the range of samples taken from a normal population. *Biometrika*, **17**, 364.

THE FISHER-YATES TEST OF SIGNIFICANCE IN 2×2 CONTINGENCY TABLES

By D. J. FINNEY
Lecturer in the Design and Analysis of Scientific Experiment, University of Oxford

1. Introduction

One of the tests of significance most frequently required in applications of statistics to biological problems is that for the 2×2 contingency table. Briefly, the problem may be typified by the following statement: Given two series of observations, with individual results assumed independent of one another and classified as either 'success' or 'failure', do the proportions of successes observed in the two series differ more widely than might reasonably be expected if the population values of these proportions are equal? If the expected numbers of successes and failures, calculated from the null hypothesis, are all moderately large, a simple and well-known form of χ^2 test may be used; this test gives a good approximation to the true probability of a deviation from equality as great as that observed, especially if the adjustment usually termed 'Yates's correction' is applied (Yates, 1934; Fisher, 1946, §§21, 21·01).

Not infrequently, however, the expected numbers are too small for the χ^2 test to be trusted. Fisher has recommended that, if any one of the expectations is less than 5, special steps should be taken; Aitken (1944) and Cramér (1946) suggest a minimum of 10. Investigations by Cochran (1942) indicate that no simple rule of this kind is completely adequate, and that the magnitudes of all four expectations affect the quality of the approximation. Yates (1934) and Fisher (1946, §21·02) have given an exact method for the evaluation of the required probability, for use when small frequencies make any χ^2 approximation unreliable. Their method uses the four marginal totals as ancillary statistics: that is to say, the probability is obtained as a relative frequency amongst all configurations having the same marginal totals. More recent writers, notably Barnard (1947) and Pearson (1947), have questioned whether this assumption of fixed marginal totals is logically justifiable for all 2×2 contingency tables; Barnard has discussed the specifications of several distinct problems, and maintains that the Fisher-Yates test is applicable to only one of these. Nevertheless, the test is undoubtedly the right one for at least one important class of problem, and for this may be described as an exact method.†

The Fisher-Yates method requires rather tedious calculations, though the labour can be much reduced by use of a simple table of the logarithm of the factorial function, such as that given by Fisher & Yates (1948, Table XXX). The purpose of the present paper is to present a table of significance levels for their exact test which is applicable when all the frequencies are small. Yates (1934) and Fisher & Yates (1948, Table VIII) give a concise table which, with the aid of a small amount of calculation, leads to a good approximation to the exact probability; the new table provides tests of significance only, but has the advantage of direct entry instead of preliminary calculation. Rules for the use of the table are given in §2, in sufficient detail for the non-mathematical reader; a brief account of the method of construction follows in §3.

† The probability associated with any of the classes of problem distinguished by Barnard is satisfactorily approximated by that obtained from χ^2 when all expected frequencies are large.

2. The table of significance levels and its use

The table which follows may be used to test the significance of the deviation from proportionality in any 2×2 contingency table having both frequencies in one of its margins less than or equal to 15. The contingency table must first be put in the form

	Number of		Number of observations
	Successes	Failures	
Series I	a	$A-a$	A
Series II	b	$B-b$	B
Total	$a+b$	$A+B-a-b$	$A+B$

where Series I is defined to be that which makes $A \geqslant B$, and the type of observation conventionally regarded as a 'success' is that which makes $a/A \geqslant b/B$. For small integers, the right arrangement can usually be made on sight, especially if the second condition is used in the form

$$aB \geqslant bA.$$

Providing that $A \leqslant 15$, the table at the end of this section may then be entered in the section for A, the sub-section for B, and the line for a. If b is equal to or less than the integer in the column headed 0·05 or 0·01, a/A is significantly *greater than* b/B (single-tail test) at the probability level 0·05 or 0·01 respectively. If b is equal to or less than the integer in the column headed 0·025 or 0·005, a/A is significantly *different from* b/B (two-tail test) at the probability level 0·05 or 0·01 respectively. A dash, or absence of any entry, for some combination of A, B, and a indicates that no contingency table in that class is significant. The probability corresponding to b will generally be less than that shown at the head of the column, and the true numerical values are shown in small type. Those in the 0·025 and 0·005 columns should be doubled if used in a two-tail test.

From the mathematical point of view, the two pairs of marginal totals play equivalent roles in the test. Consequently, if only one set satisfies the condition that neither total exceeds 15, that margin may be conventionally regarded as relating to the two series, and the other classification must then be taken as 'success' and 'failure' (see example below). If both margins satisfy the condition, the two possible arrangements will lead to identical conclusions.

The use of the table presented here may be extended by noting that any contingency table more extreme than one known to show significant deviation from proportionality must itself be non-significant. Thus, even though one of each pair of marginal totals exceeds 15, a test of significance without calculation may still be possible. In particular, if the standard arrangement of the contingency table gives

a	$A-a$	A
b	$B-b$	B
$a+b$	$A+B-a-b$	$A+B$

with $A > 15 \geqslant B$ and $aB \geqslant bA$, the deviation from proportionality will be significant if

$$\begin{array}{cc|c} 15+a-A & A-a & 15 \\ b & B-b & B \\ \hline 15+a+b-A & A+B-a-b & 15+B \end{array}$$

is significant (but not necessarily non-significant otherwise), and the deviation will be non-significant if

$$\begin{array}{cc|c} a & 15-a & 15 \\ b & B-b & B \\ \hline a+b & 15+B-a-b & 15+B \end{array}$$

is not significant (but not necessarily significant otherwise).

As an example of the use of the table, Lange's data on criminality among twin brothers or sisters of criminals (Fisher, 1946, § 21·01) may be examined. The contingency table below shows the numbers of twin brothers or sisters of criminals who had also been convicted, separately for monzygotic and dizygotic (but like-sexed) twins.

	Not convicted	Convicted	Total
Dizygotic	15	2	17
Monozygotic	3	10	13
Total	18	12	30

Since $15/17 > 3/13$ the category 'not convicted' is regarded as success. Consider the contingency table

$$\begin{array}{cc|c} 13 & 2 & 15 \\ 3 & 10 & 13 \\ \hline 16 & 12 & 28 \end{array}$$

in which $A = 15$, $B = 13$, and $a = 13$. The null hypothesis is that, in the general population, freedom from conviction is equally frequent amongst dizygotic and monozygotic sibs. If the only deviation from the null hypothesis which the investigator is prepared to consider is that monozygotic twins behave more similarly than dizygotic, he will require a one-tail significance test. The table below shows 4 as the 0·01 significance level of b (with a true probability of 0·004, not 0·010). Hence $b = 3$ is significant: by the rule given, the deviation from proportionality in the original table is significant evidence that criminality is more frequent among monozygotic twins of criminals than among dizygotic twins of criminals. Fisher gives the exact probability for the original data as 1/2150. On the other hand, if the investigator were concerned only to demonstrate a significant difference between the frequencies of freedom from conviction for the two types of twin, irrespective of whether monozygotic or dizygotic should show the higher value, he would use a two-tail test; for

$A = 15$, $B = 13$, $a = 13$, the value of b in the 0·005 column is again 4 (though generally it will be less than for the one-tail test), and again the evidence against the null hypothesis is judged significant, both for the modified and for the original contingency table.

As an illustration of the interchangeability of marginal totals, the same data may be examined by arranging the contingency table as:

	Dizygotic	Monozygotic	Total
Convicted	15	3	18
Not convicted	2	10	12
Total	17	13	30

in which 'dizygotic' is classed as success. Logically the meaning may be different, but the test of significance is the same. The modified contingency table

12	3	15
2	10	12
14	13	27

has $A = 15$, $B = 12$, $a = 12$; $b = 2$ is judged significant by comparison with the tabular values 3 at 0·01 or 2 at 0·005, and therefore again significance is attained on either the one-tail or the two-tail test.

The table follows on pp. 149–54.

3. Construction of the table

Any 2×2 contingency table to which a test of significance is to be applied must first be written in the standard form shown at the beginning of § 2, in which the event regarded as a success and the arrangement of the rows are so chosen that $A \geqslant B$ and $a/A \geqslant b/B$. If $a/A = b/B$, there is no need to proceed further, as the data then accord perfectly with the null hypothesis; if $a/A > b/B$, of necessity $a > b$. On the null hypothesis that the population values for the proportions of successes are equal, and for the specified set of marginal totals, the probability of this set of frequencies is (Fisher, 1946, § 21·02):

$$P_b = P(b \mid A, B, a+b)$$
$$= \frac{A!\,B!\,(a+b)!\,(A+B-a-b)!}{(A+B)!} \times \frac{1}{a!\,b!\,(A-a)!\,(B-b)!}.$$

The first factor in this expression is dependent only on the marginal totals, the second depends upon the internal cell frequencies. The probability that a deviation from equality of the two proportions as great as or greater than that observed should occur by chance is then P_b^*, where

$$P_b^* = P^*(b \mid A, B, a+b)$$
$$= P_b + P_{b-1} + P_{b-2} + \ldots + P_k,$$

and k is the greater of the two quantities 0 and $(a+b-A)$. In the summation the four marginal totals are kept fixed. Barnard (1947) has discussed the logic of the test in detail. He distinguishes certain classes of problem for which, in his view, the method described here is inappropriate; this topic has been a source of some controversy, to which the present paper is not intended as a contribution.

Table of significance levels of b

(Values of b in bold type; corresponding probabilities, P_b^*, in small type)

	a	Probability 0·05	0·025	0·01	0·005
A=3 B=3	3	**0** ·050	—	—	—
A=4 B=4	4	**0** ·014	**0** ·014	—	—
3	4	**0** ·029	—	—	—
A=5 B=5	5	**1** ·024	**1** ·024	**0** ·004	**0** ·004
	4	**0** ·024	**0** ·024	—	—
4	5	**1** ·048	**0** ·008	**0** ·008	—
	4	**0** ·040	—	—	—
3	5	**0** ·018	**0** ·018	—	—
2	5	**0** ·048	—	—	—
A=6 B=6	6	**2** ·030	**1** ·008	**1** ·008	**0** ·001
	5	**1** ·040	**0** ·008	**0** ·008	—
	4	**0** ·030	—	—	—
5	6	**1** ·015+	**0** ·015+	**0** ·002	**0** ·002
	5	**0** ·013	**0** ·013	—	—
	4	**0** ·045+	—	—	—
4	6	**1** ·033	**0** ·005−	**0** ·005−	**0** ·005−
	5	**0** ·024	**0** ·024	—	—
3	6	**0** ·012	**0** ·012	—	—
	5	**0** ·048	—	—	—
2	6	**0** ·036	—	—	—
A=7 B=7	7	**3** ·035−	**2** ·010+	**1** ·002	**1** ·002
	6	**1** ·015−	**1** ·015−	**0** ·002	**0** ·002
	5	**0** ·010+	**0** ·010+	—	—
	4	**0** ·035−	—	—	—
6	7	**2** ·021	**2** ·021	**1** ·005−	**1** ·005−
	6	**1** ·025+	**0** ·004	**0** ·004	**0** ·004
	5	**0** ·016	**0** ·016	—	—
	4	**0** ·049	—	—	—
5	7	**2** ·045+	**1** ·010+	**0** ·001	**0** ·001
	6	**1** ·045+	**0** ·008	**0** ·008	—
	5	**0** ·027	—	—	—
4	7	**1** ·024	**1** ·024	**0** ·003	**0** ·003
	6	**0** ·015+	**0** ·015+	—	—
	5	**0** ·045+	—	—	—
3	7	**0** ·008	**0** ·008	**0** ·008	—
	6	**0** ·033	—	—	—
2	7	**0** ·028	—	—	—

	a	Probability 0·05	0·025	0·01	0·005
A=8 B=8	8	**4** ·038	**3** ·013	**2** ·003	**2** ·003
	7	**2** ·020	**2** ·020	**1** ·005+	**0** ·001
	6	**1** ·020	**1** ·020	**0** ·003	**0** ·003
	5	**0** ·013	**0** ·013	—	—
	4	**0** ·038	—	—	—
7	8	**3** ·026	**2** ·007	**2** ·007	**1** ·001
	7	**2** ·035−	**1** ·009	**1** ·009	**0** ·001
	6	**1** ·032	**0** ·006	**0** ·006	—
	5	**0** ·019	**0** ·019	—	—
6	8	**2** ·015−	**2** ·015−	**1** ·003	**1** ·003
	7	**1** ·016	**1** ·016	**0** ·002	**0** ·002
	6	**0** ·009	**0** ·009	**0** ·009	—
	5	**0** ·028	—	—	—
5	8	**2** ·035−	**1** ·007	**1** ·007	**0** ·001
	7	**1** ·032	**0** ·005−	**0** ·005−	**0** ·005−
	6	**0** ·016	**0** ·016	—	—
	5	**0** ·044	—	—	—
4	8	**1** ·018	**1** ·018	**0** ·002	**0** ·002
	7	**0** ·010+	**0** ·010+	—	—
	6	**0** ·030	—	—	—
3	8	**0** ·006	**0** ·006	**0** ·006	—
	7	**0** ·024	**0** ·024	—	—
2	8	**0** ·022	**0** ·022	—	—
A=9 B=9	9	**5** ·041	**4** ·015−	**3** ·005−	**3** ·005−
	8	**3** ·025−	**3** ·025−	**2** ·008	**1** ·002
	7	**2** ·028	**1** ·008	**1** ·008	**0** ·001
	6	**1** ·025−	**1** ·025−	**0** ·005−	**0** ·005−
	5	**0** ·015−	**0** ·015−	—	—
	4	**0** ·041	—	—	—
8	9	**4** ·029	**3** ·009	**3** ·009	**2** ·002
	8	**3** ·043	**2** ·013	**1** ·003	**1** ·003
	7	**2** ·044	**1** ·012	**0** ·002	**0** ·002
	6	**1** ·036	**0** ·007	**0** ·007	—
	5	**0** ·020	**0** ·020	—	—
7	9	**3** ·019	**3** ·019	**2** ·005−	**2** ·005−
	8	**2** ·024	**2** ·024	**1** ·006	**0** ·001
	7	**1** ·020	**1** ·020	**0** ·003	**0** ·003
	6	**0** ·010+	**0** ·010+	—	—
	5	**0** ·029	—	—	—
6	9	**3** ·044	**2** ·011	**1** ·002	**1** ·002
	8	**2** ·047	**1** ·011	**0** ·001	**0** ·001
	7	**1** ·035−	**0** ·006	**0** ·006	—
	6	**0** ·017	**0** ·017	—	—
	5	**0** ·042	—	—	—

Table of significance levels of b (continued)

	a	Probability			
		0·05	0·025	0·01	0·005
A=9 B=5	9	2 ·027	1 ·005−	1 ·005−	1 ·005−
	8	1 ·023	1 ·023	0 ·003	0 ·003
	7	0 ·010+	0 ·010+	—	—
	6	0 ·028	—	—	—
4	9	1 ·014	1 ·014	0 ·001	0 ·001
	8	0 ·007	0 ·007	0 ·007	—
	7	0 ·021	0 ·021	—	—
	6	0 ·049	—	—	—
3	9	1 ·045+	0 ·005−	0 ·005−	0 ·005−
	8	0 ·018	0 ·018	—	—
	7	0 ·045+	—	—	—
2	9	0 ·018	0 ·018	—	—
A=10 B=10	10	6 ·043	5 ·016	4 ·005+	3 ·002
	9	4 ·029	3 ·010−	3 ·010−	2 ·003
	8	3 ·035−	2 ·012	1 ·003	1 ·003
	7	2 ·035−	1 ·010−	1 ·010−	0 ·002
	6	1 ·029	0 ·005+	0 ·005+	—
	5	0 ·016	0 ·016	—	—
	4	0 ·043	—	—	—
9	10	5 ·033	4 ·011	3 ·003	3 ·003
	9	4 ·050−	3 ·017	2 ·005−	2 ·005−
	8	2 ·019	2 ·019	1 ·004	1 ·004
	7	1 ·015−	1 ·015−	0 ·002	0 ·002
	6	1 ·040	0 ·008	0 ·008	—
	5	0 ·022	0 ·022	—	—
8	10	4 ·023	4 ·023	3 ·007	2 ·002
	9	3 ·032	2 ·009	2 ·009	1 ·002
	8	2 ·031	1 ·008	1 ·008	0 ·001
	7	1 ·023	1 ·023	0 ·004	0 ·004
	6	0 ·011	0 ·011	—	—
	5	0 ·029	—	—	—
7	10	3 ·015−	3 ·015−	2 ·003	2 ·003
	9	2 ·018	2 ·018	1 ·004	1 ·004
	8	1 ·013	1 ·013	0 ·002	0 ·002
	7	1 ·036	0 ·006	0 ·006	—
	6	0 ·017	0 ·017	—	—
	5	0 ·041	—	—	—
6	10	3 ·036	2 ·008	2 ·008	1 ·001
	9	2 ·036	1 ·008	1 ·008	0 ·001
	8	1 ·024	1 ·024	0 ·003	0 ·003
	7	0 ·010+	0 ·010+	—	—
	6	0 ·026	—	—	—
5	10	2 ·022	2 ·022	1 ·004	1 ·004
	9	1 ·017	1 ·017	0 ·002	0 ·002
	8	1 ·047	0 ·007	0 ·007	—
	7	0 ·019	0 ·019	—	—
	6	0 ·042	—	—	—
A=10 B=4	10	1 ·011	1 ·011	0 ·001	0 ·001
	9	1 ·041	0 ·005−	0 ·005−	0 ·005−
	8	0 ·015−	0 ·015−	—	—
	7	0 ·035−	—	—	—
3	10	1 ·038	0 ·003	0 ·003	0 ·003
	9	0 ·014	0 ·014	—	—
	8	0 ·035−	—	—	—
2	10	0 ·015+	0 ·015+	—	—
	9	0 ·045+	—	—	—
A=11 B=11	11	7 ·045+	6 ·018	5 ·006	4 ·002
	10	5 ·032	4 ·012	3 ·004	3 ·004
	9	4 ·040	3 ·015−	2 ·004	2 ·004
	8	3 ·043	2 ·015−	1 ·004	1 ·004
	7	2 ·040	1 ·012	0 ·002	0 ·002
	6	1 ·032	0 ·006	0 ·006	—
	5	0 ·018	0 ·018	—	—
	4	0 ·045+	—	—	—
10	11	6 ·035+	5 ·012	4 ·004	4 ·004
	10	4 ·021	4 ·021	3 ·007	2 ·002
	9	3 ·024	3 ·024	2 ·007	1 ·002
	8	2 ·023	2 ·023	1 ·006	0 ·001
	7	1 ·017	1 ·017	0 ·003	0 ·003
	6	1 ·043	0 ·009	0 ·009	—
	5	0 ·023	0 ·023	—	—
9	11	5 ·026	4 ·008	4 ·008	3 ·002
	10	4 ·038	3 ·012	2 ·003	2 ·003
	9	3 ·040	2 ·012	1 ·003	1 ·003
	8	2 ·035−	1 ·009	1 ·009	0 ·001
	7	1 ·025−	1 ·025−	0 ·004	0 ·004
	6	0 ·012	0 ·012	—	—
	5	0 ·030	—	—	—
8	11	4 ·018	4 ·018	3 ·005−	3 ·005−
	10	3 ·024	3 ·024	2 ·006	1 ·001
	9	2 ·022	2 ·022	1 ·005−	1 ·005−
	8	1 ·015−	1 ·015−	0 ·002	0 ·002
	7	1 ·037	0 ·007	0 ·007	—
	6	0 ·017	0 ·017	—	—
	5	0 ·040	—	—	—
7	11	4 ·043	3 ·011	2 ·002	2 ·002
	10	3 ·047	2 ·013	1 002	1 ·002
	9	2 ·039	1 ·009	1 ·009	0 ·001
	8	1 ·025−	1 ·025−	0 ·004	0 ·004
	7	0 ·010+	0 ·010+	—	—
	6	0 ·025−	0 ·025−	—	—
6	11	3 ·029	2 ·006	2 ·006	1 ·001
	10	2 ·028	1 ·005+	1 ·005+	0 ·001
	9	1 ·018	1 ·018	0 ·002	0 ·002

Table of significance levels of b (continued)

	a	Probability 0·05	0·025	0·01	0·005
A=11 B=6	8	1 ·043	0 ·007	0 ·007	—
	7	0 ·017	0 ·017	—	—
	6	0 ·037	—	—	—
5	11	2 ·018	2 ·018	1 ·003	1 ·003
	10	1 ·013	1 ·013	0 ·001	0 ·001
	9	1 ·036	0 ·005-	0 ·005-	0 ·005-
	8	0 ·013	0 ·013	—	—
	7	0 ·029	—	—	—
4	11	1 ·009	1 ·009	1 ·009	0 ·001
	10	1 ·033	0 ·004	0 ·004	0 ·004
	9	0 ·011	0 ·011	—	—
	8	0 ·026	—	—	—
3	11	1 ·033	0 ·003	0 ·003	0 ·003
	10	0 ·011	0 ·011	—	—
	9	0 ·027	—	—	—
2	11	0 ·013	0 ·013	—	—
	10	0 ·038	—	—	—
A=12 B=12	12	8 ·047	7 ·019	6 ·007	5 ·002
	11	6 ·034	5 ·014	4 ·005-	4 ·005-
	10	5 ·045-	4 ·018	3 ·006	2 ·002
	9	4 ·050-	3 ·020	2 ·006	1 ·001
	8	3 ·050-	2 ·018	1 ·005-	1 ·005-
	7	2 ·045-	1 ·014	0 ·002	0 ·002
	6	1 ·034	0 ·007	0 ·007	—
	5	0 ·019	0 ·019	—	—
	4	0 ·047	—	—	—
11	12	7 ·037	6 ·014	5 ·005-	5 ·005-
	11	5 ·024	5 ·024	4 ·008	3 ·002
	10	4 ·029	3 ·010+	2 ·003	2 ·003
	9	3 ·030	2 ·009	2 ·009	1 ·002
	8	2 ·026	1 ·007	1 ·007	0 ·001
	7	1 ·019	1 ·019	0 ·003	0 ·003
	6	1 ·045-	0 ·009	0 ·009	—
	5	0 ·024	0 ·024	—	—
10	12	6 ·029	5 ·010-	5 ·010-	4 ·003
	11	5 ·043	4 ·015+	3 ·005-	3 ·005-
	10	4 ·048	3 ·017	2 ·005-	2 ·005-
	9	3 ·046	2 ·015-	1 ·004	1 ·004
	8	2 ·038	1 ·010+	0 ·002	0 ·002
	7	1 ·026	0 ·005-	0 ·005-	0 ·005-
	6	0 ·012	0 ·012	—	—
	5	0 ·030	—	—	—
9	12	5 ·021	5 ·021	4 ·006	3 ·002
	11	4 ·029	3 ·009	3 ·009	2 ·002
	10	3 ·029	2 ·008	2 ·008	1 ·002
	9	2 ·024	2 ·024	1 ·006	0 ·001
	8	1 ·016	1 ·016	0 ·002	0 ·002
A=12 B=9	7	1 ·037	0 ·007	0 ·007	—
	6	0 ·017	0 ·017	—	—
	5	0 ·039	—	—	—
8	12	5 ·049	4 ·014	3 ·004	3 ·004
	11	3 ·018	3 ·018	2 ·004	2 ·004
	10	2 ·015+	2 ·015+	1 ·003	1 ·003
	9	2 ·040	1 ·010-	1 ·010-	0 ·001
	8	1 ·025-	1 ·025-	0 ·004	0 ·004
	7	0 ·010+	0 ·010+	—	—
	6	0 ·024	0 ·024	—	—
7	12	4 ·036	3 ·009	3 ·009	2 ·002
	11	3 ·038	2 ·010-	2 ·010-	1 ·002
	10	2 ·029	1 ·006	1 ·006	0 ·001
	9	1 ·017	1 ·017	0 ·002	0 ·002
	8	1 ·040	0 ·007	0 ·007	—
	7	0 ·016	0 ·016	—	—
	6	0 ·034	—	—	—
6	12	3 ·025-	3 ·025-	2 ·005-	2 ·005-
	11	2 ·022	2 ·022	1 ·004	1 ·004
	10	1 ·013	1 ·013	0 ·002	0 ·002
	9	1 ·032	0 ·005-	0 ·005-	0 ·005-
	8	0 ·011	0 ·011	—	—
	7	0 ·025-	0 ·025-	—	—
	6	0 ·050-	—	—	—
5	12	2 ·015-	2 ·015-	1 ·002	1 ·002
	11	1 ·010-	1 ·010-	1 ·010-	0 ·001
	10	1 ·028	0 ·003	0 ·003	0 ·003
	9	0 ·009	0 ·009	0 ·009	—
	8	0 ·020	0 ·020	—	—
	7	0 ·041	—	—	—
4	12	2 ·050	1 ·007	1 ·007	0 ·001
	11	1 ·027	0 ·003	0 ·003	0 ·003
	10	0 ·008	0 ·008	0 ·008	—
	9	0 ·019	0 ·019	—	—
	8	0 ·038	—	—	—
3	12	1 ·029	0 ·002	0 ·002	0 ·002
	11	0 ·009	0 ·009	0 ·009	—
	10	0 ·022	0 ·022	—	—
	9	0 ·044	—	—	—
2	12	0 ·011	0 ·011	—	—
	11	0 ·033	—	—	—
A=13 B=13	13	9 ·048	8 ·020	7 ·007	6 ·003
	12	7 ·037	6 ·015+	5 ·006	4 ·002
	11	6 ·048	5 ·021	4 ·008	3 ·002
	10	4 ·024	4 ·024	3 ·008	2 ·002
	9	3 ·024	3 ·024	2 ·008	1 ·002
	8	2 ·021	2 ·021	1 ·006	0 ·001

Table of significance levels of b (continued)

	a	Probability 0·05	0·025	0·01	0·005
A=13 B=13	7	2 ·048	1 ·015+	0 ·003	0 ·003
	6	1 ·037	0 ·007	0 ·007	—
	5	0 ·020	0 ·020	—	—
	4	0 ·048	—	—	—
12	13	8 ·039	7 ·015−	6 ·005+	5 ·002
	12	6 ·027	5 ·010−	5 ·010−	4 ·003
	11	5 ·033	4 ·013	3 ·004	3 ·004
	10	4 ·036	3 ·013	2 ·004	2 ·004
	9	3 ·034	2 ·011	1 ·003	1 ·003
	8	2 ·029	1 ·008	1 ·008	0 ·001
	7	1 ·020	1 ·020	0 ·004	0 ·004
	6	1 ·046	0 ·010−	0 ·010−	—
	5	0 ·024	0 ·024	—	—
11	13	7 ·031	6 ·011	5 ·003	5 ·003
	12	6 ·048	5 ·018	4 ·006	3 ·002
	11	4 ·021	4 ·021	3 ·007	2 ·002
	10	3 ·021	3 ·021	2 ·006	1 ·001
	9	3 ·050−	2 ·017	1 ·004	1 ·004
	8	2 ·040	1 ·011	0 ·002	0 ·002
	7	1 ·027	0 ·005−	0 ·005−	0 ·005−
	6	0 ·013	0 ·013	—	—
	5	0 ·030	—	—	—
10	13	6 ·024	6 ·024	5 ·007	4 ·002
	12	5 ·035−	4 ·012	3 ·003	3 ·003
	11	4 ·037	3 ·012	2 ·003	2 ·003
	10	3 ·033	2 ·010+	1 ·002	1 ·002
	9	2 ·026	1 ·006	1 ·006	0 ·001
	8	1 ·017	1 ·017	0 ·003	0 ·003
	7	1 ·038	0 ·007	0 ·007	—
	6	0 ·017	0 ·017	—	—
	5	0 ·038	—	—	—
9	13	5 ·017	5 ·017	4 ·005−	4 ·005−
	12	4 ·023	4 ·023	3 ·007	2 ·001
	11	3 ·022	3 ·022	2 ·006	1 ·001
	10	2 ·017	2 ·017	1 ·004	1 ·004
	9	2 ·040	1 ·010+	0 ·001	0 ·001
	8	1 ·025−	1 ·025−	0 ·004	0 ·004
	7	0 ·010+	0 ·010+	—	—
	6	0 ·023	0 ·023	—	—
	5	0 ·049	—	—	—
8	13	5 ·042	4 ·012	3 ·003	3 ·003
	12	4 ·047	3 ·014	2 ·003	2 ·003
	11	3 ·041	2 ·011	1 ·002	1 ·002
	10	2 ·029	1 ·007	1 ·007	0 ·001
	9	1 ·017	1 ·017	0 ·002	0 ·002
	8	1 ·037	0 ·006	0 ·006	—
	7	0 ·015−	0 ·015−	—	—
	6	0 ·032	—	—	—
7	13	4 ·031	3 ·007	3 ·007	2 ·001
	12	3 ·031	2 ·007	2 ·007	1 ·001
A=13 B=7	11	2 ·022	2 ·022	1 ·004	1 ·004
	10	1 ·012	1 ·012	0 ·002	0 ·002
	9	1 ·029	0 ·004	0 ·004	0 ·004
	8	0 ·010+	0 ·010+	—	—
	7	0 ·022	0 ·022	—	—
	6	0 ·044	—	—	—
6	13	3 ·021	3 ·021	2 ·004	2 ·004
	12	2 ·017	2 ·017	1 ·003	1 ·003
	11	2 ·046	1 ·010−	1 ·010−	0 ·001
	10	1 ·024	1 ·024	0 ·003	0 ·003
	9	1 ·050−	0 ·008	0 ·008	—
	8	0 ·017	0 ·017	—	—
	7	0 ·034	—	—	—
5	13	2 ·012	2 ·012	1 ·002	1 ·002
	12	2 ·044	1 ·008	1 ·008	0 ·001
	11	1 ·022	1 ·022	0 ·002	0 ·002
	10	1 ·047	0 ·007	0 ·007	—
	9	0 ·015−	0 ·015−	—	—
	8	0 ·029	—	—	—
4	13	2 ·044	1 ·006	1 ·006	0 ·000
	12	1 ·022	1 ·022	0 ·002	0 ·002
	11	0 ·006	0 ·006	0 ·006	—
	10	0 ·015−	0 ·015−	—	—
	9	0 ·029	—	—	—
3	13	1 ·025	1 ·025	0 ·002	0 ·002
	12	0 ·007	0 ·007	0 ·007	—
	11	0 ·018	0 ·018	—	—
	10	0 ·036	—	—	—
2	13	0 ·010−	0 ·010−	0 ·010−	—
	12	0 ·029	—	—	—
A=14 B=14	14	10 ·049	9 ·020	8 ·008	7 ·003
	13	8 ·038	7 ·016	6 ·006	5 ·002
	12	6 ·023	6 ·023	5 ·009	4 ·003
	11	5 ·027	4 ·011	3 ·004	3 ·004
	10	4 ·028	3 ·011	2 ·003	2 ·003
	9	3 ·027	2 ·009	2 ·009	1 ·002
	8	2 ·023	2 ·023	1 ·006	0 ·001
	7	1 ·016	1 ·016	0 ·003	0 ·003
	6	1 ·038	0 ·008	0 ·008	—
	5	0 ·020	0 ·020	—	—
	4	0 ·049	—	—	—
13	14	9 ·041	8 ·016	7 ·006	6 ·002
	13	7 ·029	6 ·011	5 ·004	5 ·004
	12	6 ·037	5 ·015+	4 ·005+	3 ·002
	11	5 ·041	4 ·017	3 ·006	2 ·001
	10	4 ·041	3 ·016	2 ·005−	2 ·005−
	9	3 ·038	2 ·013	1 ·003	1 ·003
	8	2 ·031	1 ·009	1 ·009	0 ·001

Table of significance levels of b (continued)

	a	Probability			
		0·05	0·025	0·01	0·005
A=14 B=13	7	1 ·021	1 ·021	0 ·004	0 ·004
	6	1 ·048	0 ·010+	—	—
	5	0 ·025−	0 ·025−	—	—
12	14	8 ·033	7 ·012	6 ·004	6 ·004
	13	6 ·021	6 ·021	5 ·007	4 ·002
	12	5 ·025+	4 ·009	4 ·009	3 ·003
	11	4 ·026	3 ·009	3 ·009	2 ·002
	10	3 ·024	3 ·024	2 ·007	1 ·002
	9	2 ·019	2 ·019	1 ·005−	1 ·005−
	8	2 ·042	1 ·012	0 ·002	0 ·002
	7	1 ·028	0 ·005+	0 ·005+	—
	6	0 ·013	0 ·013	—	—
	5	0 ·030	—	—	—
11	14	7 ·026	6 ·009	6 ·009	5 ·003
	13	6 ·039	5 ·014	4 ·004	4 ·004
	12	5 ·043	4 ·016	3 ·005−	3 ·005−
	11	4 ·042	3 ·015−	2 ·004	2 ·004
	10	3 ·036	2 ·011	1 ·003	1 ·003
	9	2 ·027	1 ·007	1 ·007	0 ·001
	8	1 ·017	1 ·017	0 ·003	0 ·003
	7	1 ·038	0 ·007	0 ·007	—
	6	0 ·017	0 ·017	—	—
	5	0 ·038	—	—	—
10	14	6 ·020	6 ·020	5 ·006	4 ·002
	13	5 ·028	4 ·009	4 ·009	3 ·002
	12	4 ·028	3 ·009	3 ·009	2 ·002
	11	3 ·024	3 ·024	2 ·007	1 ·001
	10	2 ·018	2 ·018	1 ·004	1 ·004
	9	2 ·040	1 ·011	0 ·002	0 ·002
	8	1 ·024	1 ·024	0 ·004	0 ·004
	7	0 ·010−	0 ·010−	0 ·010−	—
	6	0 ·022	0 ·022	—	—
	5	0 ·047	—	—	—
9	14	6 ·047	5 ·014	4 ·004	4 ·004
	13	4 ·018	4 ·018	3 ·005−	3 ·005−
	12	3 ·017	3 ·017	2 ·004	2 ·004
	11	3 ·042	2 ·012	1 ·002	1 ·002
	10	2 ·029	1 ·007	1 ·007	0 ·001
	9	1 ·017	1 ·017	0 ·002	0 ·002
	8	1 ·036	0 ·006	0 ·006	—
	7	0 ·014	0 ·014	—	—
	6	0 ·030	—	—	—
8	14	5 ·036	4 ·010−	4 ·010−	3 ·002
	13	4 ·039	3 ·011	2 ·002	2 ·002
	12	3 ·032	2 ·008	2 ·008	1 ·001
	11	2 ·022	2 ·022	1 ·005−	1 ·005−
	10	2 ·048	1 ·012	0 ·002	0 ·002
	9	1 ·026	0 ·004	0 ·004	0 ·004
	8	0 ·009	0 ·009	0 ·009	—
	7	0 ·020	0 ·020	—	—
	6	0 ·040	—	—	—
A=14 B=7	14	4 ·026	3 ·006	3 ·006	2 ·001
	13	3 ·025	2 ·006	2 ·006	1 ·001
	12	2 ·017	2 ·017	1 ·003	1 ·003
	11	2 ·041	1 ·009	1 ·009	0 ·001
	10	1 ·021	1 ·021	0 ·003	0 ·003
	9	1 ·043	0 ·007	0 ·007	—
	8	0 ·015−	0 ·015−	—	—
	7	0 ·030	—	—	—
6	14	3 ·018	3 ·018	2 ·003	2 ·003
	13	2 ·014	2 ·014	1 ·002	1 ·002
	12	2 ·037	1 ·007	1 ·007	0 ·001
	11	1 ·018	1 ·018	0 ·002	0 ·002
	10	1 ·038	0 ·005+	0 ·005+	—
	9	0 ·012	0 ·012	—	—
	8	0 ·024	0 ·024	—	—
	7	0 ·044	—	—	—
5	14	2 ·010+	2 ·010+	1 ·001	1 ·001
	13	2 ·037	1 ·006	1 ·006	0 ·001
	12	1 ·017	1 ·017	0 ·002	0 ·002
	11	1 ·038	0 ·005−	0 ·005−	0 ·005−
	10	0 ·011	0 ·011	—	—
	9	0 ·022	0 ·022	—	—
	8	0 ·040	—	—	—
4	14	2 ·039	1 ·005−	1 ·005−	1 ·005−
	13	1 ·019	1 ·019	0 ·002	0 ·002
	12	1 ·044	0 ·005−	0 ·005−	0 ·005−
	11	0 ·011	0 ·011	—	—
	10	0 ·023	0 ·023	—	—
	9	0 ·041	—	—	—
3	14	1 ·022	1 ·022	0 ·001	0 ·001
	13	0 ·006	0 ·006	0 ·006	—
	12	0 ·015−	0 ·015−	—	—
	11	0 ·029	—	—	—
2	14	0 ·008	0 ·008	0 ·008	—
	13	0 ·025	0 ·025	—	—
	12	0 ·050	—	—	—
A=15 B=15	15	11 ·050−	10 ·021	9 ·008	8 ·003
	14	9 ·040	8 ·018	7 ·007	6 ·003
	13	7 ·025+	6 ·010+	5 ·004	5 ·004
	12	6 ·030	5 ·013	4 ·005−	4 ·005−
	11	5 ·033	4 ·013	3 ·005−	3 ·005−
	10	4 ·033	3 ·013	2 ·004	2 ·004
	9	3 ·030	2 ·010+	1 ·003	1 ·003
	8	2 ·025+	1 ·007	1 ·007	0 ·001
	7	1 ·018	1 ·018	0 ·003	0 ·003
	6	1 ·040	0 ·008	0 ·008	—
	5	0 ·021	0 ·021	—	—
	4	0 ·050−	—	—	—

Table of significance levels of b (continued)

	a	Probability 0·05	0·025	0·01	0·005
	15	10 ·042	9 ·017	8 ·006	7 ·002
	14	8 ·031	7 ·013	6 ·005−	6 ·005−
	13	7 ·041	6 ·017	5 ·007	4 ·002
	12	6 ·046	5 ·020	4 ·007	3 ·002
	11	5 ·048	4 ·020	3 ·007	2 ·002
	10	4 ·046	3 ·018	2 ·006	1 ·001
	9	3 ·041	2 ·014	1 ·004	1 ·004
	8	2 ·033	1 ·009	1 ·009	0 ·001
	7	1 ·022	1 ·022	0 ·004	0 ·004
	6	1 ·049	0 ·011	—	—
	5	0 ·025+	—	—	—
13	15	9 ·035−	8 ·013	7 ·005−	7 ·005−
	14	7 ·023	7 ·023	6 ·009	5 ·003
	13	6 ·029	5 ·011	4 ·004	4 ·004
	12	5 ·031	4 ·012	3 ·004	3 ·004
	11	4 ·030	3 ·011	2 ·003	2 ·003
	10	3 ·026	2 ·008	2 ·008	1 ·002
	9	2 ·020	2 ·020	1 ·005+	0 ·001
	8	2 ·043	1 ·013	0 ·002	0 ·002
	7	1 ·029	0 ·005+	0 ·005+	—
	6	0 ·013	0 ·013	—	—
	5	0 ·031	—	—	—
12	15	8 ·028	7 ·010−	7 ·010−	6 ·003
	14	7 ·043	6 ·016	5 ·006	4 ·002
	13	6 ·049	5 ·019	4 ·007	3 ·002
	12	5 ·049	4 ·019	3 ·006	2 ·002
	11	4 ·045+	3 ·017	2 ·005−	2 ·005−
	10	3 ·038	2 ·012	1 ·003	1 ·003
	9	2 ·028	1 ·007	1 ·007	0 ·001
	8	1 ·018	1 ·018	0 ·003	0 ·003
	7	1 ·038	0 ·007	0 ·007	—
	6	0 ·017	0 ·017	—	—
	5	0 ·037	—	—	—
11	15	7 ·022	7 ·022	6 ·007	5 ·002
	14	6 ·032	5 ·011	4 ·003	4 ·003
	13	5 ·034	4 ·012	3 ·003	3 ·003
	12	4 ·032	3 ·010+	2 ·003	2 ·003
	11	3 ·026	2 ·008	2 ·008	1 002
	10	2 ·019	2 ·019	1 ·004	1 ·004
	9	2 ·040	1 ·011	0 ·002	0 ·002
	8	1 ·024	1 ·024	0 ·004	0 ·004
	7	1 ·049	0 ·010−	0 ·010−	—
	6	0 ·022	0 ·022	—	—
	5	0 ·046	—	—	—
10	15	6 ·017	6 ·017	5 ·005−	5 ·005−
	14	5 ·023	5 ·023	4 ·007	3 ·002
	13	4 ·022	4 ·022	3 ·007	2 ·001
	12	3 ·018	3 ·018	2 ·005−	2 ·005−
	11	3 ·042	2 ·013	1 ·003	1 ·003
	10	2 ·029	1 ·007	1 ·007	0 ·001
	9	1 ·016	1 ·016	0 ·002	0 ·002
	8	1 ·034	0 ·006	0 ·006	—
	7	0 ·013	0 ·013	—	—
	6	0 ·028	—	—	—
9	15	6 ·042	5 ·012	4 ·003	4 ·003
	14	5 ·047	4 ·015−	3 ·004	3 ·004
A=15 B=9	13	4 ·042	3 ·013	2 ·003	2 ·003
	12	3 ·032	2 ·009	2 ·009	1 ·002
	11	2 ·021	2 ·021	1 ·005−	1 ·005−
	10	2 ·045−	1 ·011	0 ·002	0 ·002
	9	1 ·024	1 ·024	0 ·004	0 ·004
	8	1 ·048	0 ·009	0 ·009	—
	7	0 ·019	0 ·019	—	—
	6	0 ·037	—	—	—
8	15	5 ·032	4 ·008	4 ·008	3 ·002
	14	4 ·033	3 ·009	3 ·009	2 ·002
	13	3 ·026	2 ·006	2 ·006	1 ·001
	12	2 ·017	2 ·017	1 ·003	1 ·003
	11	2 ·037	1 ·008	1 ·008	0 ·001
	10	1 ·019	1 ·019	0 ·003	0 ·003
	9	1 ·038	0 ·006	0 ·006	—
	8	0 ·013	0 ·013	—	—
	7	0 ·026	—	—	—
	6	0 ·050−	—	—	—
7	15	4 ·023	4 ·023	3 ·005−	3 ·005−
	14	3 ·021	3 ·021	2 ·004	2 ·004
	13	2 ·014	2 ·014	1 ·002	1 ·002
	12	2 ·032	1 ·007	1 ·007	0 ·001
	11	1 ·015+	1 ·015+	0 ·002	0 ·002
	10	1 ·032	0 ·005−	0 ·005−	0 ·005−
	9	0 ·010+	0 ·010+	—	—
	8	0 ·020	0 ·020	—	—
	7	0 ·038	—	—	—
6	15	3 ·015+	3 ·015+	2 ·003	2 ·003
	14	2 ·011	2 ·011	1 ·002	1 ·002
	13	2 ·031	1 ·006	1 ·006	0 ·001
	12	1 ·014	1 ·014	0 ·002	0 ·002
	11	1 ·029	0 ·004	0 ·004	0 ·004
	10	0 ·009	0 ·009	0 ·009	—
	9	0 ·017	0 ·017	—	—
	8	0 ·032	—	—	—
5	15	2 ·009	2 ·009	2 ·009	1 ·001
	14	2 ·032	1 ·005−	1 ·005−	1 ·005−
	13	1 ·014	1 ·014	0 ·001	0 ·001
	12	1 ·031	0 ·004	0 ·004	0 ·004
	11	0 ·008	0 ·008	0 ·008	—
	10	0 ·016	0 ·016	—	—
	9	0 ·030	—	—	—
4	15	2 ·035+	1 ·004	1 ·004	1 ·004
	14	1 ·016	1 ·016	0 ·001	0 ·001
	13	1 ·037	0 ·004	0 ·004	0 ·004
	12	0 ·009	0 ·009	0 ·009	—
	11	0 ·018	0 ·018	—	—
	10	0 ·033	—	—	—
3	15	1 ·020	1 ·020	0 ·001	0 ·001
	14	0 ·005−	0 ·005−	0 ·005−	0 ·005−
	13	0 ·012	0 ·012	—	—
	12	0 ·025−	0 ·025−	—	—
	11	0 ·043	—	—	—
2	15	0 ·007	0 ·007	0 ·007	—
	14	0 ·022	0 ·022	—	—
	13	0 ·044	—	—	—

The table given in § 2 enables tests of significance, at probability levels of 0·05, 0·025, 0·01 and 0·005, to be made by direct reference for any 2×2 contingency table having $B \leqslant A \leqslant 15$. The section of the table for any particular pair of values A, B was constructed by giving to $n = a+b$ successively all integral values from $(A+B)$ down to zero; for each n, b was given in turn the values k, $k+1$, $k+2$, ... (where k is the greater of the two quantities 0 and $n-A$), P_b was formed and the cumulative sum P_b^* was recorded. The calculations with any set of A, B, n were stopped as soon as P_b^* exceeded 0·05 or would obviously exceed 0·05 for the next higher value of b. The calculations were tedious, but by no means as lengthy as might appear from this account, for after a little experience extreme values of n, which would always give $P_b^* > 0{\cdot}05$, could be ignored, and trends in P_b^* often enabled values of b which would exceed the limit to be foreseen without calculation of P_b. Of course, in any case of doubt, P_b^* was evaluated.

The probabilities were determined to five places of decimals, but any full publication of these would occupy a great deal of space. For many practical purposes, all that is required is a test of significance at an arbitrary level of probability. This can be conveniently made by use of a table showing, for each combination of A, B, and a, the greatest value of b for which P_b^* lies beyond the chosen significance level. The values of b required can be seen immediately from systematic inspection of the calculations just described. For example, the calculations show the following 2×2 configurations with their associated values of P_b^*.

8	3	11
3	7	10
11	10	21

$P_b^* = 0{\cdot}0635$

8	3	11
2	8	10
10	11	21

$P_b^* = 0{\cdot}0226$

8	3	11
1	9	10
9	12	21

$P_b^* = 0{\cdot}0058$

8	3	11
0	10	10
8	13	21

$P_b^* = 0{\cdot}0008$

Consequently

$$P^*(b \mid 11, 10, 11) \leqslant 0{\cdot}05 \quad \text{when} \quad b \leqslant 2,$$
$$P^*(b \mid 11, 10, 10) \leqslant 0{\cdot}025 \quad \text{when} \quad b \leqslant 2,$$
$$P^*(b \mid 11, 10, 9) \leqslant 0{\cdot}01 \quad \text{when} \quad b \leqslant 1,$$

and

$$P^*(b \mid 11, 10, 8) \leqslant 0{\cdot}005 \quad \text{when} \quad b = 0.$$

These values of b, 2, 2, 1, and 0, are therefore tabulated in § 2 under the appropriate probabilities. The P_b^* corresponding to each tabulated b, which is usually less than the significance level, is shown in small type; for any smaller b, P_b^* will be even lower. From the existing calculations, of course, significance levels for any other selected probability less than 0·05 could easily be read, but 0·05, 0·025, 0·01 and 0·005 seem sufficient for the present.

The standard tables of the χ^2 distribution (Fisher & Yates, 1948, Table IV), provide a two-tail test when applied to a 2×2 contingency table. That is to say, on the null hypothesis that the proportion of successes in the two populations compared are equal, and assuming that frequencies are sufficiently large for the sampling distributions of the proportions to be taken as normal, the test is based on the probability of obtaining a difference as great as or greater than $(a/A - b/B)$ *in either direction* under conditions of random sampling. If a single-tail test is wanted, it can be obtained simply by entering the χ^2 table in the column corresponding to twice the level of significance. For small frequencies, however, the discrete nature of the distribution for configurations with fixed marginal frequencies cannot be ignored, and unless either $A = B$ or $2n = (A+B)$ this distribution is not symmetrical. The exact test of significance described at the beginning of this section is a single-tail test, being

based on the probability of a deviation from proportionality as great as or greater than that observed and in the direction of the observed deviation. In general, no deviation in the opposite direction will have exactly the same probability. For example, for the configuration

$$\begin{array}{cc|c} 10 & 0 & 10 \\ 4 & 5 & 9 \\ \hline 14 & 5 & 19 \end{array}$$

the probability is $$P^* = P = \frac{10!\,9!\,14!\,5!}{19!\,10!\,4!\,5!} = 0{\cdot}0108.$$

No deviation in the opposite direction can be regarded as the equivalent of this either in the sense of having the same difference in observed proportions or in the sense of having the same probability; the extreme configuration at the other tail is

$$\begin{array}{cc|c} 5 & 5 & 10 \\ 9 & 0 & 9 \\ \hline 14 & 5 & 19 \end{array}$$

for which $$P^* = P = 0{\cdot}0217.$$

If a two-tail test is wanted when the frequencies are small, a new convention must first be introduced. The only satisfactory procedure, which is consistent with the practice for larger frequencies (when the χ^2 distribution is used), is to regard the two-tail probability as given by $2P^*$. In other words, the two-tail significance tests corresponding to single-tail tests at the conventional probability levels 0·05 and 0·01 are given by comparing P^* with 0·025 and 0·005 respectively. It is this consideration which governed the choice of probability levels for tabulation in § 2.

4. Summary

This paper contains a table from which a test of significance of the deviation from proportionality in a 2 × 2 contingency table, based upon the Fisher-Yates exact probability method, can be read directly. Any 2 × 2 contingency table for which neither member of one pair of marginal totals exceeds 15 can be tested in this way, and a simple rule extends the applicability of the table to certain contingency tables with larger frequencies. The method of construction of the table is described.

I am indebted to a number of past and present members of my staff for the extensive and tedious calculations on which the table is based, especial thanks being due to Miss M. Callow.

REFERENCES

Aitken, A. C. (1944). *Statistical Mathematics*, 3rd ed. Edinburgh: Oliver and Boyd.

Barnard, G. A. (1947). Significance tests for 2 × 2 tables. *Biometrika*, **34**, 123.

Cochran, W. G. (1942). The χ^2 correction for continuity. *Iowa St. Coll. J. Sci.* **16**, 421.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.

Fisher, R. A. (1946). *Statistical Methods for Research Workers*, 10th ed. Edinburgh: Oliver and Boyd.

Fisher, R. A. & Yates, F. (1948). *Statistical Tables for Biological, Agricultural and Medical Research*, 3rd ed. Edinburgh: Oliver and Boyd.

Pearson, E. S. (1947). The choice of statistical tests illustrated on the interpretation of data classed in a 2 × 2 table. *Biometrika*, **34**, 139.

Yates, F. (1934). Contingency tables involving small numbers and the χ^2 test. *J. Roy. Statist. Soc.* Suppl. **1**, 217.

THE POWER FUNCTION OF THE TEST FOR THE DIFFERENCE BETWEEN TWO PROPORTIONS IN A 2×2 TABLE

By P. B. PATNAIK, *University College, London*

1. Introduction

Neyman & Pearson's (1933) conception of the power of a test of a statistical hypothesis, H_0, was developed, in the first instance as a means of guiding the choice between alternative tests. This, it was shown, could be done by comparing the effectiveness of the tests in discriminating between H_0 and a set of admissible alternative hypotheses regarded as most relevant to the question under test. Where there is no doubt about the most appropriate test and no sequential scheme of sampling is possible, the power function may play a useful part in indicating, before the data are collected, how large the samples should be to avoid an inconclusive result. If this procedure is to be easily applied, a ready means must be available of calculating the power of the test for a given significance level and sample size. The tables of the power function of the t-test (Neyman & Tokarska, 1936) and Tang's Tables (1938) applicable in the Analysis of Variance, are examples of such aids. The present paper aims at providing in simple, if approximate, form a means of determining the power function of the test for the difference between two proportions.

The test may be briefly outlined as follows. In two 'infinite' populations the proportions of individuals possessing a character A are $p_1(A)$ and $p_2(A)$ respectively. Random samples of m and n are drawn from the two populations and the result is represented in a 2×2 table, thus:

Table 1. *Number of individuals*

	With A	Without A	Total
Sample from first population	a	c	m
Sample from second population	b	d	n
Total	r	s	N

We may then wish to apply the test in two forms:

(i) The 'two-sided' form. To test H_0 that $p_1 = p_2$, bearing in mind the two-sided alternatives that $p_1 < p_2$ and $p_1 > p_2$.

(ii) The 'one-sided' form. To test H_0 that $p_1 \leqslant p_2$, bearing in mind the one-sided alternatives $p_1 > p_2$. These will be referred to in the following sections as cases (i) and (ii).

In obtaining a solution we shall regard the sample space as two-dimensioned as in Pearson's (1947) Problem II or Barnard's (1947) '2×2 comparative trial'. Fig. 1 (a) roughly illustrates the sample space for the case $m = n = 50$. A possible sampling result is represented by a point (a, b) in this lattice of $51 \times 51 = 2601$ points. Sample points having $a+b = r$ are

said to lie on the 'diagonal' r of the lattice. In the case of small samples, or even for large samples when (a, b) falls near the $(0, 0)$ or (m, n) corners of the lattice, considerable difficulties arise owing to the discontinuity of the distribution. We shall be concerned in the first place with examining the case where it is justifiable to apply the test by referring to the normal probability scale the ratio

$$\left.\begin{aligned} u &= \frac{a - \dfrac{rm}{N}}{\sqrt{\dfrac{mnrs}{N^2(N-1)}}} \\ &= \frac{\dfrac{a}{m} - \dfrac{b}{n}}{\sqrt{\left[\dfrac{r}{N}\left(1-\dfrac{r}{N}\right)\left(\dfrac{1}{m}+\dfrac{1}{n}\right)\right]}} = \frac{(ad-bc)\sqrt{N}}{\sqrt{(mnrs)}} \quad \text{(approx.)}, \end{aligned}\right\} \tag{1}$$

where the second and third alternative forms are equivalent to the first if the factor $N-1$ is replaced by N. The first form brings out the fact that u is the ratio of the deviation from the mean to the standard deviation, in the hypergeometric series representing the conditional distribution of a for r fixed. The second form arises from the classical approach in which a difference of observed proportions is compared with an estimate of its standard error. The third, when squared, gives the commonly quoted form of the expression for χ^2, with one degree of freedom, in a 2×2 table.

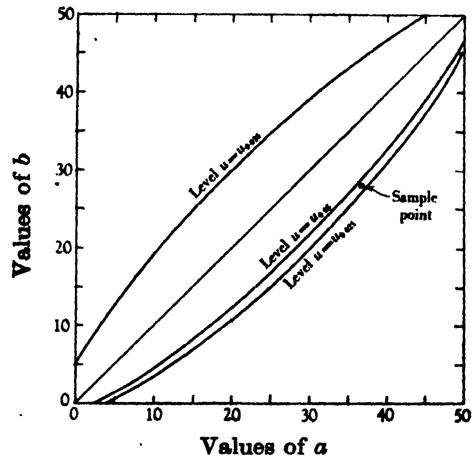


Fig. 1 (*a*). Sample space ($m = n = 50$).

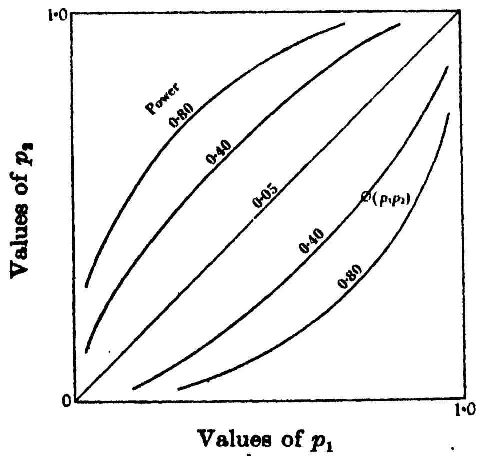


Fig. 1 (*b*). Power contours ($m = n = 50$).

Write u_α for the 100α percentage point of the normal distribution $N(0, 1)$ i.e.

$$\int_{u_\alpha}^{\infty} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}u^2} = \alpha,$$

then using the two-sided test (i), we should reject the hypothesis that $p_1 = p_2$ at the significance level α when $|u| > u_{\frac{1}{2}\alpha}$; in the case of the one-sided test, we should reject the hypothesis that $p_1 \leqslant p_2$ at the same significance level when $u > u_\alpha$. Thus, in the example illustrated in Fig. 1 (*a*), it is seen from the position of the sample point, that H_0 would not be rejected at the 5 % level if the test is in form (i), but would be rejected at the 5 % level if it is in form (ii).

The boundaries of the critical region associated with the test, taking say, case (i), are formed by obtaining from the equation

$$\frac{a-\dfrac{mr}{N}}{\sqrt{\dfrac{mnrs}{N^2(N-1)}}} = u_{\frac{1}{2}\alpha}$$

the 'cut-off' points $(a, r-a)$ on each diagonal r in the lattice and joining them as in Fig. 1 (*a*). Subject to the error involved in the approximation, the chance is α that the sample point falls in the critical region when p_1 and p_2 have a common, though unknown, value i.e. in Neyman & Pearson's notation, $P\{E \epsilon w_\alpha \mid p_1 = p_2\} = \alpha$. If in fact $p_1 \neq p_2$, then the power of the test of H_0 with regard to the alternative hypothesis $H_1(p_1, p_2)$ is the chance that the point (a, b) falls in the critical region when sampling from populations with proportions p_1 and p_2. Or, formally,

$$P\{E \epsilon w_\alpha \mid p_1, p_2\} = P\{\mid u \mid > u_{\frac{1}{2}\alpha} \mid p_1, p_2\} \quad \text{for case (i),}$$

or

$$= P\{u > u_\alpha \mid p_1, p_2\} \quad \text{for case (ii).}$$

This is the total probability density at all the discrete points (a, b) included in the critical region. The problem is to express this in a readily calculable form.

If this is done, two types of application are evident:

(1) When the decision has been made to take two samples of, say, 50, or when the available data happen to consist in samples of this size, we may ask, 'what is the chance that the test described will show a difference in observed proportions significant at the 5 % level when in fact, p_1 and p_2 are as different, say, as 0·50 and 0·65?'

(2) On the other hand, we may use the theory to ask in advance how large the samples should be so that the risk of failing to detect a given difference* between p_1 and p_2 which is considered to be of importance, shall be acceptably small. For example, we may ask, 'what sizes of samples should we take so that in applying our test we may have a high, say a 90 %, chance of detecting that the proportions are not equal when, in fact, they are as different as 0·50 and 0·65?'

It is clear that for given m, n and α the power of the test will be constant on certain contours in the p_1, p_2 space such as those shown in Fig. 1 (*b*). We shall examine the approximate form of these contours and show that in the important case when $m = n = \frac{1}{2}N$, the family of contours is independent of N and α, although the power associated with a particular contour will be a function of N and α, which has been tabled. Throughout the investigation, approximations are made of the type involved in representing binomial or hypergeometric distributions by normal distributions. The adequacy of these approximations is examined in certain cases.

2. EXACT VALUES FOR THE POWER FUNCTION OF THE TEST BASED ON THE RATIO u

Since a and b are independent, it follows that for the population proportions, p_1 and p_2

$$p(a,b) = \frac{m!}{a!\,c!}p_1^a q_1^c \times \frac{n!}{b!\,d!}p_2^b q_2^d \quad (q_1 = 1-p_1,\ q_2 = 1-p_2). \tag{2}$$

If our test consists in rejecting $H_0(p_1 = p_2)$ when $\mid u \mid > u_{\frac{1}{2}\alpha}$, its power with respect to an alternative $H_1(p_1 \neq p_2)$ is the sum of the values of $p(a, b)$ for all points (a, b) lying in the critical region. Exact calculation is laborious, since it involves the multiplication of terms of one binomial by the sums of terms of the other, which may be obtained from the tables of the

* Neyman & Pearson have termed this the error of the second kind.

Incomplete Beta Function. In the special case $m = 18, n = 12$, discussed in Pearson's paper, exact values of the power function have been calculated for $p_1, p_2 = 0\cdot1(0\cdot1)\,0\cdot9$ for the two-sided test and for significance levels 10 and 2 %; they are tabulated in Tables 2(*a*) and

Table 2(*a*). *Power function for case* $m = 18, n = 12$. *Significance level* $\alpha = 0\cdot10$

Approximate values are shown in parentheses

p_2 \ p_1	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9
0·9	1·000 (1·00)	0·997 (0·999)	0·975 (0·992)	0·907 (0·96)	0·773 (0·87)	0·574 (0·70)	0·356 (0·44)	0·170 (0·21)	0·088 (0·10)
0·8	0·997 (0·999)	0·972 (0·980)	0·882 (0·91)	0·732 (0·77)	0·494 (0·56)	0·289 (0·34)	0·144 (0·17)	0·094 (0·10)	0·202
0·7	0·977 (0·987)	0·880 (0·90)	0·722 (0·73)	0·487 (0·51)	0·272 (0·30)	0·136 (0·15)	0·092 (0·10)	0·159	0·393
0·6	0·918 (0·94)	0·742 (0·75)	0·515 (0·51)	0·287 (0·29)	0·142 (0·15)	0·099 (0·10)	0·155	0·320	0·610
0·5	0·796 (0·83)	0·534 (0·54)	0·309 (0·30)	0·156 (0·15)	0·100 (0·10)	0·156	0·309	0·534	0·796
0·4	0·610 (0·64)	0·320 (0·32)	0·155 (0·15)	0·099 (0·10)	0·142	0·287	0·515	0·742	0·918
0·3	0·393 (0·40)	0·159 (0·16)	0·092 (0·10)	0·136	0·272	0·487	0·722	0·880	0·977
0·2	0·202 (0·20)	0·094 (0·10)	0·144	0·289	0·494	0·732	0·882	0·972	0·997
0·1	0·088 (0·10)	0·170	0·356	0·574	0·773	0·907	0·975	0·997	1·000

Table 2(*b*). *Power function for case* $m = 18$, $n = 12$. *Significance level* $\alpha = 0\cdot02$

Approximate values are shown in parentheses

p_2 \ p_1	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9
0·9	0·998 (1·00)	0·976 (0·993)	0·902 (0·96)	0·752 (0·86)	0·532 (0·68)	0·290 (0·43)	0·103 (0·20)	0·017 (0·06)	0·006 (0·02)
0·8	0·961 (0·990)	0·882 (0·91)	0·694 (0·75)	0·490 (0·52)	0·265 (0·30)	0·115 (0·12)	0·035 (0·05)	0·016 (0·02)	0·055
0·7	0·909 (0·94)	0·714 (0·73)	0·460 (0·48)	0·248 (0·26)	0·111 (0·11)	0·041 (0·04)	0·021 (0·02)	0·050	0·177
0·6	0·767 (0·81)	0·500 (0·49)	0·262 (0·25)	0·111 (0·11)	0·041 (0·04)	0·023 (0·02)	0·048	0·138	0·361
0·5	0·625 (0·66)	0·293 (0·28)	0·124 (0·11)	0·045 (0·04)	0·022 (0·02)	0·045	0·124	0·293	0·625
0·4	0·361 (0·37)	0·138 (0·13)	0·048 (0·04)	0·023 (0·02)	0·041	0·111	0·262	0·500	0·767
0·3	0·177 (0·18)	0·050 (0·05)	0·021 (0·02)	0·041	0·111	0·248	0·460	0·714	0·909
0·2	0·055 (0·06)	0·016 (0·02)	0·035	0·115	0·265	0·490	0·694	0·882	0·961
0·1	0·006 (0·02)	0·017	0·103	0·290	0·532	0·752	0·902	0·976	0·998

2(*b*).* When $p_1 = p_2$ the value of the power function should reduce to α. But in these tables we see that the values on the diagonal, $p_1 = p_2$ are not exactly equal to 0·10 or 0·02. The discrepancy is due to the fact that continuous approximations have been made to the discontinuous distributions in the formulation of the test.

Interpolation between the calculated values in Table 2(*a*), leads to the power contours shown in Fig. 2. The test has the same power of establishing significance when sampling from any populations for which p_1, p_2 lies on a given contour. The chances of establishing significance are written alongside the contours. The surface for which the ordinates at p_1, p_2 are equal to the power, may be called the power surface of the test.

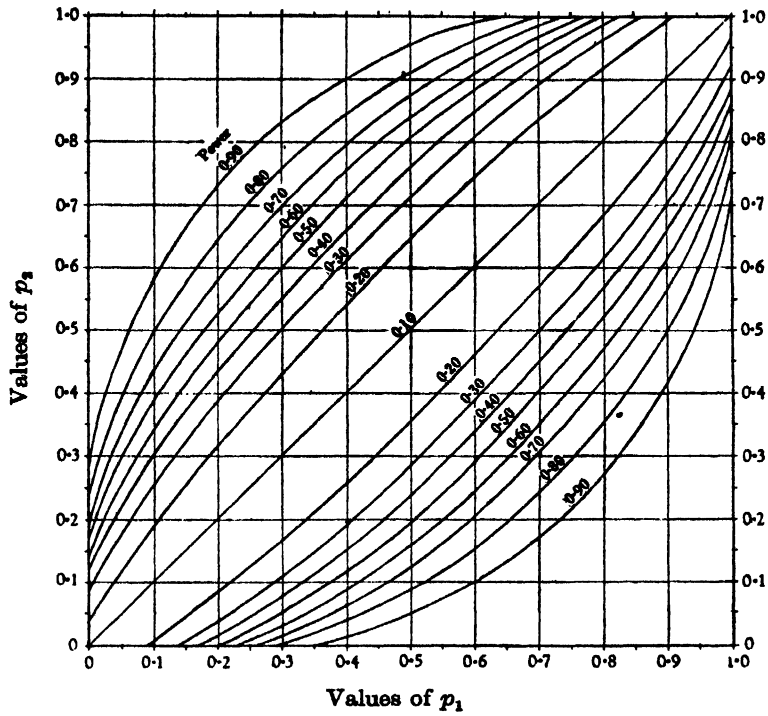


Fig. 2. Power contours (special case $m = 18$, $n = 12$) for significance level 0·10.

We can see from the contour diagram that with samples of this size (18 and 12), even when the p's are as different as 0·6 and 0·3 there is only a 50 % chance of establishing significance. If instead of the high level $\alpha = 0{\cdot}10$ employed here, a lower level were chosen, then the p's must differ even more to give the same chance of establishing significance. Thus the diagram illustrates well how inadequate small samples are to establish what would ordinarily be regarded as a difference of some importance.

It can also be seen that $|p_1 - p_2|$ is nearly constant on a power contour near the middle of the square; e.g. the difference is roughly 0·3 on the 0·5 contour within the range (0·45, 0·15) and (0·75, 0·45).

* For these calculations, I am indebted to the members of the Statistics Department of University College, London, in particular to Mr V. D. Gangolli.

3. Approximations to the power function

We will now consider an approximation to the distribution of a under the hypothesis H_1, under which the population proportions are p_1 and p_2. From (1)

$$p(a,r) = \frac{N!}{r!\,s!} p_2^r q_2^s \left(\frac{q_1}{q_2}\right)^m \times \left\{\frac{m!\,n!\,r!\,s!}{N!\,a!\,b!\,c!\,d!}\right\} \left(\frac{p_1 q_2}{p_2 q_1}\right)^a.$$

If we replace the hypergeometric term in the curled brackets by the ordinate of a normal curve having the mean and S.D. of the hypergeometric series, then

$$p(a,r) = \frac{N!}{r!\,s!} p_2^r q_2^s \left(\frac{q_1}{q_2}\right)^m \times \frac{1}{\sqrt{(2\pi)}\sqrt{\frac{mnrs}{N^2(N-1)}}} \exp\left[-\frac{\left(a-\frac{rm}{N}\right)^2}{2\frac{mnrs}{N^2(N-1)}}\right] \times \left(\frac{p_1 q_2}{p_2 q_1}\right)^a.$$

Writing $\left(\frac{p_1 q_2}{p_2 q_1}\right)^a$ as $\exp\left[a \log_e \left(\frac{p_1 q_2}{p_2 q_1}\right)\right]$, collecting the terms containing a and making a perfect square, we obtain $\quad p(a,r) = p(r) \times p(a \mid r),$

where
$$p(r) = \frac{N!}{r!\,s!} p_2^r q_2^s \left(\frac{q_1}{q_2}\right)^m \exp\left[\frac{rm}{N} \log_e \frac{p_1 q_2}{p_2 q_1} + \frac{mnrs}{2N^2(N-1)} \left(\log_e \frac{p_1 q_2}{p_2 q_1}\right)^2\right], \tag{3}$$

and
$$p(a|r) = \frac{1}{\sqrt{(2\pi)}\sqrt{\frac{mnrs}{N^2(N-1)}}} \exp\left[-\frac{\left(a-\frac{rm}{N}-\frac{mnrs}{N^2(N-1)}\log_e \frac{p_1 q_2}{p_2 q_1}\right)^2}{2\frac{mnrs}{N^2(N-1)}}\right]. \tag{4}$$

Thus (4) is the approximate conditional distribution of a on the diagonal $r = a+b$ and is seen to be normal, with

$$\text{mean} = \frac{rm}{N} + \frac{mnrs}{N^2(N-1)} \log_e \frac{p_1 q_2}{p_2 q_1}$$

and
$$\text{S.D.} = \sqrt{\frac{mnrs}{N^2(N-1)}}.$$

Defining
$$u = \frac{a - \frac{rm}{N}}{\sqrt{\frac{mnrs}{N^2(N-1)}}}$$

as in (1), equation (4) becomes

$$p(u|r) = \frac{1}{\sqrt{(2\pi)}} \exp\left[-\frac{1}{2}\left(u - \sqrt{\frac{mnrs}{N^2(N-1)}} \log_e \frac{p_1 q_2}{p_2 q_1}\right)^2\right]$$
$$= \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}(u-h(r))^2}, \tag{5}$$

where
$$h(r) = \sqrt{\frac{mnrs}{N^2(N-1)}} \log_e \frac{p_1 q_2}{p_2 q_1}, \tag{6}$$

and is a function of r only, since $s(= N-r)$ and the other quantities are given.

If p_1 and p_2 are equal, then (5) reduces to the distribution of u under H_0,

$$p(u) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}u^2}, \tag{7}$$

which is the normal approximation used in obtaining the test criterion. This distribution is independent of r; it is this fact that makes the critical regions on each diagonal r similar regions which can be combined into a single region in the two-dimensioned sample space.

But the distribution of u under H_1, given by (5) is not independent of r. It is normal, with the same S.D. as for (7), i.e. unity, but with its mean shifted by $h(r)$.

What may be termed the 'conditional power', for r fixed, with regard to $H_1(p_1 \neq p_2)$ is then for case (i),

$$P(|u| > u_{\frac{1}{2}\alpha} \mid r) = \int_{-\infty}^{-u_{\frac{1}{2}\alpha}} p(u|r)\,du + \int_{u_{\frac{1}{2}\alpha}}^{\infty} p(u|r)\,du = 1 - \frac{1}{\sqrt{(2\pi)}}\int_{-u_{\frac{1}{2}\alpha}-h(r)}^{u_{\frac{1}{2}\alpha}-h(r)} e^{-\frac{1}{2}u^2}du, \tag{8}$$

and for case (ii), $$P(u > u_\alpha | r) = \int_{u_\alpha}^{\infty} p(u|r)\,du = \frac{1}{\sqrt{(2\pi)}}\int_{u_\alpha - h(r)}^{\infty} e^{-\frac{1}{2}u^2}\,du. \tag{9}$$

Since $p(u) = p(u|r)\,p(r)$, the 'over-all' power function or, simply, the power function is

$$\int_{-\infty}^{\infty} P(|u| > u_{\frac{1}{2}\alpha} \mid r)\,p(r)\,dr \tag{10}$$

for case (i) with a similar expression for case (ii). It is clear that even with the simplified expression (8) for the conditional power function, depending on $h(r)$, the labour involved in calculating the over-all power (10) would in general be prohibitive. Some approximation is therefore required for $p(r)$, i.e. the expression in (3). The simplest approximation is obtained by assuming r to be normally distributed. Since a and b are distributed binomially with means mp_1, and np_2 and S.D.'s mp_1q_1 and np_2q_2, $r = a+b$ can be considered as distributed normally with mean $= mp_1 + np_2$ and S.D. $= \sqrt{(mp_1q_1 + np_2q_2)}$. That is,

$$p(r) = \frac{1}{\sqrt{(2\pi)}\sqrt{(mp_1q_1+np_2q_2)}}\exp\left[-\frac{(r-(mp_1+np_2))^2}{2(mp_1q_1+np_2q_2)}\right]. \tag{11}$$

Hence, the expression (10) for the over-all power function becomes

$$\frac{1}{\sqrt{(2\pi)}\sqrt{(mp_1q_1+np_2q_2)}}\int_{-\infty}^{\infty} P(|u| > u_{\frac{1}{2}\alpha} \mid r) \times \exp\left[-\frac{1}{2}\frac{(r-(mp_1+np_2))^2}{mp_1q_1+np_2q_2}\right]dr \tag{12}$$

for case (i) with a similar expression for case (ii).

Though problems may arise where the conditional power function would be useful, clearly it is the value of the over-all power that will help in determining in advance of the experimental result how large the difference between p_1 and p_2 must be for the standard test to have a given chance of establishing significance. For in such a preliminary survey, r cannot be regarded as fixed.

4. EVALUATION OF OVER-ALL POWER

The over-all power is a function of p_1 and p_2, for given m, n and α and may be written as $\beta(p_1, p_2 \mid m, n, \alpha)$. Similarly, the conditional power function may be written as

$$\beta(p_1, p_2 \mid m, n, \alpha, r).$$

They will be denoted here for simplicity by β and $\beta(r)$ respectively.

Suppose $\mu_1 = mp_1 + np_2$, $\sigma^2 = \mu_2 = mp_1q_1 + np_2q_2$ and μ_3, μ_4, etc. equal the higher moments of the normal distribution (11). Then for case (i) or case (ii), from (12)

$$\beta = \frac{1}{\sqrt{(2\pi)}\,\sigma}\int_{-\infty}^{\infty}\exp\left[-\frac{(r-\mu_1)^2}{2\sigma^2}\right]\beta(r)\,dr.$$

Expanding $\beta(r)$ by Taylor's Theorem,

$$\beta(r) = \beta(\mu_1) + (r-\mu_1)\,\beta'(\mu_1) + \frac{(r-\mu_1)^2}{2!}\,\beta''(\mu_1) + \dots \tag{13}$$

and substituting in above, we obtain

$$\beta = \beta(\mu_1)\frac{1}{\sqrt{(2\pi)}\,\sigma}\int_{-\infty}^{\infty}\exp\left[-\frac{(r-\mu_1)^2}{2\sigma^2}\right]dr$$

$$+\beta'(\mu_1)\frac{1}{\sqrt{(2\pi)}\,\sigma}\int_{-\infty}^{\infty}\exp\left[-\frac{(r-\mu_1)^2}{2\sigma^2}\right](r-\mu_1)\,dr$$

$$+\frac{\beta''(\mu_1)}{2!}\frac{1}{\sqrt{(2\pi)}\,\sigma}\int_{-\infty}^{\infty}\exp\left[-\frac{(r-\mu_1)^2}{2\sigma^2}\right](r-\mu_1)^2\,dr+\dots$$

$$=\beta(\mu_1)+\frac{\beta''(\mu_1)}{2!}\mu_2+\frac{\beta^{(4)}(\mu_1)}{4!}\mu_4+\dots. \qquad (14)$$

It follows that a first approximation to the over-all power β is

$$\beta(\mu_1) = 1-\frac{1}{\sqrt{(2\pi)}}\int_{-u_{\frac{1}{2}\alpha}-h(\mu_1)}^{u_{\frac{1}{2}\alpha}-h(\mu_1)} e^{-\frac{1}{2}u^2}\,du \quad \text{for case (i)} \qquad (15)$$

or

$$= \frac{1}{\sqrt{(2\pi)}}\int_{u_\alpha-h(\mu_1)}^{\infty} e^{-\frac{1}{2}u^2}\,du \quad \text{for case (ii)} \qquad (16)$$

substituting μ_1 for r in the expressions (8) and (9).

A second approximation will be derived and discussed later in section 10.

5. COMPARISON OF EXACT AND APPROXIMATE VALUES OF THE POWER FUNCTION

Taking the two-sided test, its over-all power has been considered in three special cases. Using the first approximation, the values of β have been calculated for $m = 18, n = 12$ and are shown in parentheses below the exact values in Tables 2(*a*) and 2(*b*). The exact and the approximate values have also been calculated for $m = n = 15$ and $m = n = 30$ for a few combinations of p_1, p_2, selected so as to give high power which, as we shall see later, is in the range we are most interested in. These are given in Tables 3(*a*) and 3(*b*). In the latter case, since N is large, $N-1$ is replaced by N in the ratio u.

Table 3(*a*). *Showing the power of the two-sided test.* $m = n = 15$

		Significance level, α					
		0·10			0·02		
p_1	p_2	Exact value	First approx.	Second approx.	Exact value	First approx.	Second approx.
0·3	0·4	0·141	0·16	0·154	0·034	0·04	0·041
0·6	0·8	0·306	0·34	0·334	0·112	0·14	0·133
0·1	0·3	0·389	0·44	0·422	0·149	0·20	0·193
0·2	0·7	0·896	0·92	0·912	0·680	0·76	0·750
0·05	0·5	0·916	0·97	0·964	0·736	0·90	0·876
0·1	0·6	0·919	0·96	0·953	0·739	0·86	0·844
0·2	0·8	0·974	0·98	0·982	0·872	0·93	0·923
0·1	0·7	0·980	0·992	0·991	0·894	0·96	0·954

Table 3 (*b*). *Showing the power of the two-sided test.* $m = n = 30$

p_1	p_2	Significance level, α					
		0·10			0·02		
		Exact value	First approx.	Second approx.	Exact value	First approx.	Second approx.
0·05	0·3	0·884	0·93	0·904	0·631	0·78	0·752
0·1	0·4	0·885	0·91	0·903	0·691	0·75	0·736
0·3	0·7	0·937	0·95	0·947	0·807	0·83	0·824
0·2	0·6	0·945	0·96	0·957	0·839	0·86	0·852
0·1	0·5	0·977	0·988	0·985	0·902	0·94	0·934
0·2	0·7	0·993	0·996	0·996	0·965	0·98	0·974

From these tables it can be seen generally that (1) the first approximation over-estimates the power, (2) the agreement is better with large sample sizes, (3) the discrepancy is less when p_1 and p_2 are near 0·5 than when one or both is very small, i.e. when we are in a corner of the p_1, p_2-space, (4) the approximation is better for $\alpha = 0{\cdot}10$ than for $\alpha = 0{\cdot}02$.

Mathematically, the approximation is of course not very accurate even with two samples of 30. But it must be remembered (*a*) that what is needed in practice is a simple procedure for determining quickly the chance of establishing significance, and (*b*) that errors of approximation are not everywhere of the same importance. When the chance of detecting a worthwhile difference between p_1 and p_2 is only 0·40 it is clear that the sample size is inadequate and this would still be our conclusion if the approximation gave 0·45. It is perhaps only when the power approaches 0·90 that an error of this order becomes serious. If the true chance were 0·90 (odds of 9 to 1) and the approximation gave 0·95 (odds of 19 to 1) the result becomes somewhat misleading on this basis. Examination of Tables 3 (*a*) and 2 (*b*) suggests that if one value of p is likely to be less than 0·1 or greater than 0·9, the approximation is failing us for $n < 30$. But even here, as shown on p. 170 below when estimating the sample size needed to provide a given power, using Tables 4 and 5, we shall not be far out.

6. Case of equal sample sizes

Suppose $m = n = \frac{1}{2}N$. Then $\mu_1 = n(p_1 + p_2)$. Substituting in (6) and replacing $N^2(N-1)$ by $N^3 (= 8n^3)$, the error being negligible when n is not too small, we find

$$h(\mu_1) = \sqrt{n}\frac{\sqrt{2}}{4}\sqrt{[(p_1+p_2)(2-p_1-p_2)]}\log_e\frac{p_1(1-p_2)}{p_2(1-p_1)}.$$

In the case of the one-sided test, where the alternative is $p_1 > p_2$, $h(\mu_1)$ is positive. In the other case with alternatives $p_1 < p_2$ or $p_1 > p_2$, $h(\mu_1)$ is negative or positive. But from the expression (15) for the approximate power, it is seen that the sign of $h(\mu_1)$ is immaterial. So, putting

$$h = |\, h(\mu_1)\,|,$$

$$k = k(p_1, p_2) = \frac{\sqrt{2}}{4}\sqrt{[(p_1+p_2)(2-p_1-p_2)]}\left|\log_e\frac{p_1(1-p_2)}{p_2(1-p_1)}\right|, \tag{17}$$

we have

$$h = k\sqrt{n}. \tag{18}$$

From (15) and (16) it follows that to the first approximation, for given $n = \frac{1}{2}N$ and α, the power, β, is a function of k only, i.e.

$$\beta(p_1, p_2 \mid n, n, \alpha) = \beta(k \mid n, \alpha).$$

These contours of constant k have been drawn in Fig. 3 and the values of k are written alongside. Thus for $m = n$ the values of p_1, p_2 and the power of the test may be linked up as follows:

(i) Fig. 3 relates p_1, p_2 to k.

(ii) Equation (18) gives h in terms of k and n.

(iii) The normal integrals (15) and (16) give the power in terms of h and the significance level α employed in the test.

It will be seen that to the first approximation the composite hypothesis,

$$H_0(p_1 = p_2 = p, \text{ unknown}),$$

is reduced to a simple hypothesis, $k = 0$. So the test in this case may be regarded as the test of the hypothesis, $k = 0$, with alternatives, $k > 0$.

7. Tabulation

(1) *Tables of k as an alternative to Fig.* 3. From (17), k has been calculated for

$$p_1, p_2 = 0{\cdot}05(0{\cdot}05)\,0{\cdot}95$$

and is given in Table 4 (printed at the end of the paper, p. 174). Since $k(1-p_1, 1-p_2)$ has the same value as $k(p_1, p_2)$ the figures in the upper part of the table ($p_2 > p_1$) are not printed.

(2) *Table 5. The power as a function of* $h = k\sqrt{n}$ *and* α. For the two-sided test we require the integral (15). Denoting this by P, i.e.,

$$P = 1 - \frac{1}{\sqrt{(2\pi)}} \int_{-u_{\frac{1}{2}\alpha}-h}^{u_{\frac{1}{2}\alpha}-h} e^{-\frac{1}{2}u^2}\, du, \tag{19}$$

the value of P is given in columns 2–5 of Table 5 (printed at the end of the paper, p. 175). It has been calculated from *Tables of Probability Functions*, Vol. II (Federal Works Agency, New York), using Lagrangian four-point interpolation, for the levels

$$\alpha = 0{\cdot}10, 0{\cdot}05, 0{\cdot}02 \text{ and } 0{\cdot}01 \quad \text{and for} \quad h = 0{\cdot}1(0{\cdot}1)\,3{\cdot}0(0{\cdot}2)\,5{\cdot}0.$$

As h increases, the contribution to this integral from one tail rapidly becomes negligible, so that (19) approximates to

$$\frac{1}{\sqrt{(2\pi)}} \int_{u_{\frac{1}{2}\alpha}-h}^{\infty} e^{-\frac{1}{2}u^2}\, du.$$

From (16), we see that this integral is the power of the one-sided test, applied at the significance level $\frac{1}{2}\alpha$. The points at which this result is obtained to a given level of accuracy are shown in Table 5; at least two-decimal accuracy occurs below the mark *, three-decimal accuracy below †, and four below ‡.

To facilitate the use, values of $k = h/\sqrt{n}$ have been tabulated in the right-hand side of Table 5 corresponding to the h values in the first column and for

$$n = 10(5)\,50(10)\,100, 150.$$

Linear interpolation is adequate for Tables 4 and 5, except in the corners of Table 4 where Lagrangian four-point interpolation may be necessary.

8. APPLICATION OF THE POWER FUNCTION AND USE OF TABLES

Illustration 1. Laurence & Newell, while experimenting with the composition of soil composts, recorded the following results of a germination trial with *Primula sinensis* seeds (quoted in *Statistical Analysis in Biology* by K. Mather, 1946, p. 193). Two equal groups of seeds were allowed to germinate in dishes containing filter papers soaked respectively in rain water and in water allowed to seep through loam before use.

Table 6

	Germinated	Ungerminated	Total
Loam water	37	13	50
Rain water	32	18	50
Total	69	31	100

To test if the type of water affects germination, u has been calculated to be 1·11 and referred to the normal probability scale. It is seen that there is no significant difference at the 5 % level.

It might be asked what magnitude of difference could we hope to detect using two samples of 50. Suppose, for example, that for these populations 80 %, say, of seeds will germinate in loam water and 60 % in rain water; what would have been the chance of establishing significance at 5 % level?

To obtain this, we find from Fig. 3 or from Table 4 the value of k for $p_1 = 0{\cdot}8$ and $p_2 = 0{\cdot}6$. From Table 4 it is seen to be 0·318. Then we enter Table 5 in the column of $n = 50$ and find that this value of k lies between the tabulated values, 0·311 and 0·325. They correspond to the figures 0·5949 and 0·6331 in the column of P for $\alpha = 0{\cdot}05$. So, the first approximation to the power lies between these values and by linear interpolation is found to be 0·61. If the level chosen for the test is $\alpha = 0{\cdot}01$, we find that the power is only 0·37. Clearly this indicates that with two samples of 50, there is a very considerable risk of failing to establish significance when the difference in chances of germination is of this order.

Suppose now we asked how large the samples should have been to give a chance of 0·9 of establishing significance when the true percentages germinating are 80 in loam water and 60 in rain water? We then proceed as follows: In Table 5, entering the column of P under $\alpha = 0{\cdot}05$, we find that 0·90 lies between 0·8925 and 0·9251 and following the rows of these figures we see that $k = 0{\cdot}318$ lies between the figures in the columns of $n = 100$ and $n = 150$. As the interval is too wide for interpolation we find h in column 1 corresponding to $P = 0{\cdot}90$ in column 3 and then from the relation, $n = h^2/k^2$, we obtain n to be nearly 105. If the level chosen is $\alpha = 0{\cdot}01$, the samples should be of size 150 to give the same power.

Illustration 2. In the early stages of production of an important piece of electrical equipment, it has been found that the percentage of units failing under test varies, according to the batch, between 30 and 40 %. An adjustment is suggested which would be considered worthwhile if it leads to a 50 % reduction in failures. A trial is planned in which $N = 2n$ units are selected at random from a batch, half are adjusted and half not and the whole are then tested. We want to know in advance about how large should n be so that the odds are

19 to 1 that we shall not reach an inconclusive result (at the 5 % level) if there has been a 50 % reduction in failures.

Here $p_1 = 0{\cdot}30$ to $0{\cdot}40$ and $p_2 = 0{\cdot}15$ to $0{\cdot}20$. From Fig. 3 we see that the contour, $k = 0{\cdot}3$ roughly passes through the range of these points (p_1, p_2). From Table 5 we find $h = 3{\cdot}3$ for $P = 0{\cdot}95$ and $\alpha = 0{\cdot}10$ (since the test is of the one-sided form). So, $n = h^2/k^2 = 120$, nearly, a surprisingly large number.

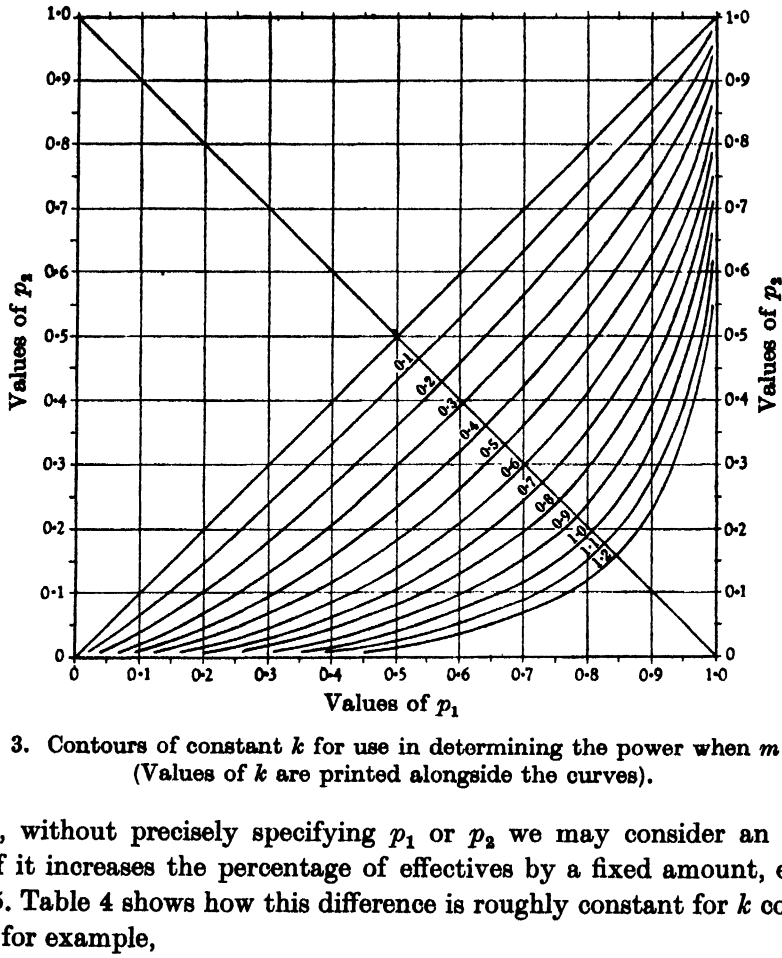


Fig. 3. Contours of constant k for use in determining the power when $m=n$. (Values of k are printed alongside the curves).

Sometimes, without precisely specifying p_1 or p_2 we may consider an improvement worthwhile if it increases the percentage of effectives by a fixed amount, e.g. such that $p_2 - p_1 = 0{\cdot}25$. Table 4 shows how this difference is roughly constant for k constant in the central area; for example,

p_1	p_2	k
0·20	0·45	0·393
0·25	0·50	0·376
0·35	0·60	0·362
0·45	0·70	0·366
0·55	0·80	0·393

That is, the contour $k = 0{\cdot}38$ passes very closely through these points (p_1, p_2). With this value of k we obtain n as before.

Illustration 3. If we have a random variable x following a distribution law which is approximately normal, the most efficient estimator of the population mean, μ, is the sample

mean $\bar{x}$. On the assumption that the variance is not changing appreciably, we should use the t-test to determine whether there has been a change in μ between the drawing of a first and second sample. Practical requirements sometimes make it preferable, or even necessary, to observe only the number of individuals in the two samples, say a and b respectively, for which x falls below a fixed level x_0. We can then test whether there has been a change in μ with the help of the ratio u of equation (1), p_1 and p_2 being the chance that $x \leqslant x_0$ for the first and second samples, respectively.

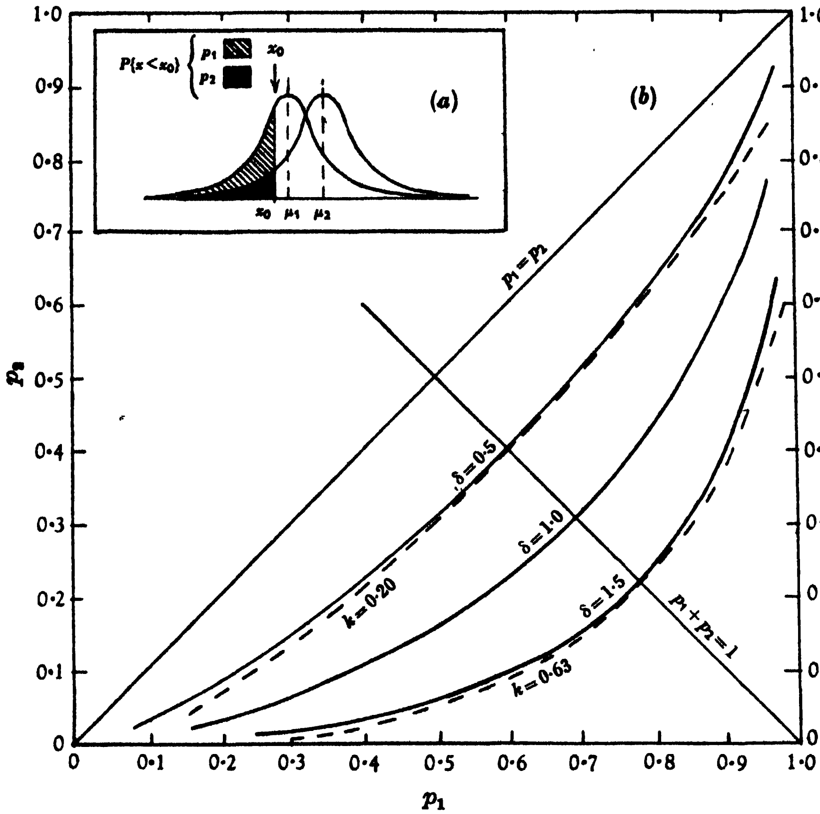


Fig. 4 (a), (b).

Examples of this problem occur:

(i) In firing two types of shot, with striking velocity x_0, against a standard proof plate and observing the numbers, a and b, that fail to perforate.

(ii) In dosage-mortality problems, where two drugs are compared only at a single dosage level, x_0.

The value of x_0 will frequently be at our choice and the generally accepted principle is to take a value so that $P\{x \leqslant x_0\}$ is in the neighbourhood of 0·5. It is possible to confirm the soundness of this procedure in terms of the power function of the u (or χ^2) test. Fig. 4(a) represents two normal distributions with means μ_1 and μ_2 and a common standard deviation, σ. p_1 and p_2 are the proportionate areas under the curves below $x = x_0$. For a given relative shift in mean, $\delta = (\mu_2 - \mu_1)/\sigma$, values of p_1 and p_2 as functions of x_0 are readily found from tables of the normal probability integral. Fig. 4(b) shows, as solid line curves, the locus of points p_1, p_2 for which $\delta = 0{\cdot}5$, 1·0 and 1·5. Our problem is to determine, for given δ, the

value of x_0 which will maximize the chance that the u-test will establish significance. We know that in large samples of equal size ($m = n$), the power of the test is constant on the contours of constant k (equation (17)). Two of these contours, for $k = 0{\cdot}20$ and $0{\cdot}63$, approximately, are shown as dotted lines in Fig. 4(b); they touch the curves for $\delta = 0{\cdot}5$ and $1{\cdot}5$, respectively, at the points where $p_1+p_2 = 1$, and elsewhere fall beyond them, as shown.

Since for small samples, we have found that the k-contours rather overestimate the power when p_1 or p_2 approach 0 or 1, it follows that the true power contour which touches a δ-curve where $p_1+p_2 = 1$, will fall even further outside it, towards the corners of the diagram, than the k-contours, as drawn. It appears therefore that:

(i) For a given δ, the usual test for a difference in proportions has a maximum chance of establishing a difference if $x_0 = \frac{1}{2}(\mu_1+\mu_2)$ and this is so for all levels of the power function.*

(ii) When $m = n$ are large and the true power contours become exactly those of k = constant, shown in Figs. 3 and 4(b), p_1+p_2 may differ considerably from 1·0 without appreciably reducing the power of the test.

9. The error in estimating the sample size necessary to ensure a given power

In section 5 it has been seen that the power obtained by the first approximation is generally an over-estimate of the true value. The effect of this will be to underestimate n in carrying out the procedure illustrated in the previous example. It is possible to determine the magnitude of this error in the neighbourhood of $n = 15$ and 30, where the exact values of the power have been found and are given in Tables 3(a) and (b). Some results of this comparison are shown in Table 7.

Table 7. *Comparing the estimates with the true values of the sample size, n*

$m = n$	p_1	p_2	True power*		k from Table 4	h for true power from Table 5		Estimate of $n = h^2/k^2$	
			For $\alpha = 0{\cdot}10$	For $\alpha = 0{\cdot}02$		$\alpha = 0{\cdot}10$	$\alpha = 0{\cdot}02$	$\alpha = 0{\cdot}10$	$\alpha = 0{\cdot}02$
15	0·05	0·5	0·916	0·736	0·930	3·03	2·96	11	11
	0·1	0·6	0·919	0·739	0·878	3·05	2·97	13	12
	0·1	0·7	0·980	0·894	1·055	3·71	3·58	13	12
30	0·05	0·3	0·884	0·631	0·563	2·84	2·66	26	23
	0·1	0·5	0·977	0·902	0·712	3·65	3·62	27	26
	0·2	0·7	0·993	0·965	0·786	4·12	4·15	28	28

* The power of the two-sided test.

For example, if $p_1 = 0{\cdot}1$, $p_2 = 0{\cdot}6$, Table 3(a) gives the true power for $m = n = 15$ as (i) 0·919 using the test with a 10 % significance level and (ii) 0·739 using the test with a 2 % level. Suppose now that we were to use Tables 4 and 5 to estimate how large the sample

* For $m = n = 30$, Table 5 shows that for $k = 0{\cdot}20$ and for a significance level $\alpha = 0{\cdot}05$, the two-sided test has a power of a little under 0·20 and for $k = 0{\cdot}63$ of a little under 0·93.

must be to give chances of 0·919 and 0·739 of establishing significance at the 10 and 2 % levels respectively, when $p_1 = 0{\cdot}1$, $p_2 = 0{\cdot}6$. From Table 4 we find that $k = 0{\cdot}878$ and interpolating in the second and third columns of Table 5 we obtain h and so n, thus:

P	α	h	$n = h^2/k^2$
0·919	0·10	3·05	12·1 or 13, taking the next higher integer
0·739	0·02	2·97	11·4 or 12, taking the next higher integer

The Table shows the extent of the underestimate of n. The error will become relatively smaller as n is increased; but some adjustment is clearly desirable.

10. A second approximation to over-all power

The first term in the right-hand side of (14) has been taken as the first approximation to the over-all power. The other terms decrease fast since the higher derivatives, $\beta''(\mu_1), \beta^{(4)}(\mu_1), \ldots$ rapidly approach zero, and a good approximation may be obtained by taking a few of these terms. We will, however, derive a second approximation by a slightly different approach.

The method of approximate product-integration developed by R. E. Beard (1947) could be employed to express β as a weighted sum of terms $\beta(r_n)$ where $r_1, r_2, \ldots, r_n$ are n values which might or might not be fixed beforehand. Considering only three such levels of r, we write formally

$$\int_{-\infty}^{\infty} \beta(r)\, p(r)\, dr = \{a_1\beta(r_1) + a_2\beta(r_2) + a_3\beta(r_3)\}\int_{-\infty}^{\infty} p(r)\, dr. \tag{20}$$

Expand the functions $\beta(r)$, $\beta(r_1)$, $\beta(r_2)$ and $\beta(r_3)$ by Taylor's Theorem to six terms as in (13) and write the remainder terms after the sixth. Then identifying the coefficients of

$$\beta(\mu_1), \quad \beta'(\mu_1), \quad \ldots, \quad \beta^{(5)}(\mu_1)$$

on both sides of (20), we have the equations

$$a_1 + a_2 + a_3 = 1$$
$$a_1(r_1 - \mu_1) + a_2(r_2 - \mu_1) + a_3(r_3 - \mu_1) = 0$$
$$a_1(r_1 - \mu_1)^2 + a_2(r_2 - \mu_1)^2 + a_3(r_3 - \mu_1)^2 = \mu_2$$

$$a_1(r_1 - \mu_1)^5 + a_2(r_2 - \mu_1)^5 + a_3(r_3 - \mu_1)^5 = 0.$$

Expressing the higher moments of the normal distribution $p(r)$ in terms of μ_2, these equations yield the solution: $a_1 = a_3 = \frac{1}{6}, \quad a_2 = \frac{2}{3}$.

$$r_1 = \mu_1 - \sqrt{(3\mu_2)}, \quad r_2 = \mu_1, \quad r_3 = \mu_1 + \sqrt{(3\mu_2)}.$$

Since the left-hand side of (20) is β, it follows that

$$\beta = \tfrac{1}{6}\beta[\mu_1 - \sqrt{(3\mu_2)}] + \tfrac{2}{3}\beta(\mu_1) + \tfrac{1}{6}\beta[\mu_1 + \sqrt{(3\mu_2)}] + R, \tag{21}$$

where R depends on the four remainder terms of the expansions $\beta(r), \ldots, \beta(r_3)$. It can be shown that

$$R = \frac{\beta^{(6)}(\xi)}{6!}\left(\mu_6 - \frac{\mu_4^2}{\mu_2}\right) = \frac{\beta^{(6)}(\xi)}{6!}\, 6\mu_2^3 \quad (-\infty < \xi < \infty).$$

We may now compare the expression for β in (21), with that in (14). Expanding the two functions $\beta[\mu_1 - \sqrt{(3\mu_2)}]$ and $\beta[\mu_1 + \sqrt{(3\mu_2)}]$ by Taylor's Theorem, we see that the right-hand side of (21), without R, includes the terms $\beta(\mu_1)$, $\dfrac{\beta''(\mu_1)}{2!}\mu_2$, $\dfrac{\beta^{(4)}(\mu_1)}{4!}\mu_4$ of (14) completely

and the following terms partially. This is also seen by comparing the form of R with the form of the remainder term of (14) after the sixth, $R_6 = \beta^{(6)}(\xi)\mu_6/6!$. Clearly,

$$\beta = \tfrac{1}{6}\beta[\mu_1 - \sqrt{(3\mu_2)}] + \tfrac{2}{3}\beta(\mu_1) + \tfrac{1}{6}\beta[\mu_1 + \sqrt{(3\mu_2)}] \tag{22}$$

gives a better approximation than

$$\beta = \beta(\mu_1) + \frac{\beta''(\mu_1)}{2!}\mu_2 + \frac{\beta^{(4)}(\mu_1)}{4!}\mu_4.$$

As it is also easier to calculate, we will regard (22) as our second approximation to the over-all power function.

We can get a similar formula with equal weights, namely

$$\beta = \frac{1}{3}\left\{\beta\left[\mu_1 - \frac{\sqrt{(3\mu_2)}}{2}\right] + \beta(\mu_1) + \beta\left[\mu_1 + \frac{\sqrt{(3\mu_2)}}{2}\right]\right\}.$$

Unlike (22), this is derived by using only the first three moments of the distribution of r. More generally, using only the first three moments, we have

$$\beta = \frac{1}{2c^2}\{\beta[\mu_1 - c\sqrt{(\mu_2)}] + \beta[\mu_1 + c\sqrt{(\mu_2)}]\} + \left(1 - \frac{1}{c^2}\right)\beta(\mu_1).$$

where c can be chosen as we like, subject to the restriction that none of the arguments should become negative.

In section 8 we have seen how the over-all power for equal samples could be obtained to the first approximation, $\beta(\mu_1)$. For the second approximation (22) the values of $\beta[\mu_1 + \sqrt{(3\mu_2)}]$ and $\beta[\mu_1 - \sqrt{(3\mu_2)}]$ may be obtained in the same manner by entering Table 5 with the values of $h[\mu_1 + \sqrt{(3\mu_2)}]$ and $h[\mu_1 - \sqrt{(3\mu_2)}]$. From (6), with $m = n$, we have

$$h[\mu_1 + \sqrt{(3\mu_2)}] = \sqrt{\frac{1}{8n}\{(n(p_1+p_2) + \sqrt{[3n(p_1q_1+p_2q_2)]})\,(2n - np_1 + p_2) - \sqrt{[3n(p_1q_1+p_2q_2)]}\}\log_e\frac{p_1q_2}{p_2q_1}}.$$

Putting $h_1 = |\,h[\mu_1 + \sqrt{(3\mu_2)}]\,|$, this becomes

$$h_1 = k\sqrt{\left\{1 + \frac{2(1-p_1-p_2)\sqrt{[3n(p_1q_1+p_2q_2)]} - 3n(p_1q_1+p_2q_2)}{n(p_1+p_2)(2-p_1-p_2)}\right\}},$$

where k is the expression (17). Similarly if

$$h_2 = |\,h[\mu_1 - \sqrt{(3\mu_2)}]\,|,$$

then
$$h_2 = k\sqrt{\left\{1 + \frac{-2(1-p_1-p_2)\sqrt{[3n(p_1q_1+p_2q_2)]} - 3n(p_1q_1+p_2q_2)}{n(p_1+p_2)(2-p_1-p_2)}\right\}}.$$

h_1 and h_2 can be calculated by obtaining k from Table 4 and substituting the values of k, p_1, p_2 and n in these two expressions.

If P_1, P_2 are the tabulated values of the integral P of (19), corresponding to h_1, h_2 and if P is the value corresponding to $h(\mu_1)$, then the second approximation is

$$\beta = \tfrac{1}{6}(P_1 + P_2 + 4P). \tag{23}$$

The value of n obtained from Table 5 as described in section 8, may be improved with the help of the second approximation. Taking this value of n we calculate h_1 and h_2 and obtain the corresponding P_1 and P_2. Then from (23),

$$P = \tfrac{3}{2}\beta - \tfrac{1}{4}(P_1 + P_2).$$

With this P we enter Table 5 again and find the improved n.

The values of the over-all power have been calculated on this second approximation for certain cases and shown alongside the first approximation values in Tables 3(*a*) and 3(*b*). It is seen that the second approximation improves the first, although it does not remove all the error.

11. General case of unequal sample sizes

The first approximation to the over-all power depends on

$$h(\mu_1) = \sqrt{\left\{\frac{mn(mp_1+np_2)(N-mp_1-np_2)}{N^3}\right\}}\log_e\frac{p_1q_2}{p_2q_1}, \tag{24}$$

taking N^3 instead of $N^2(N-1)$ in (6).

Suppose $\theta = m/n$ then

$$h = |h(\mu_1)| = \sqrt{\frac{N}{2}}\sqrt{\left\{\frac{2\theta}{(1+\theta)^4}(\theta p_1+p_2)(1+\theta\,\overline{1-p_1-p_2})\right\}}\left|\log_e\frac{p_1q_2}{p_2q_1}\right|$$

$$= \sqrt{(\tfrac{1}{2}N)}\,k(\theta), \quad \text{say.} \tag{25}$$

$k(\theta)$ corresponds to k in the case of equal samples, but in general it is a function of θ as well as p_1 and p_2. If θ is known, as for example, when m and n are given, the k function defines as before a family of contours of constant power in the p_1p_2 plane.

To obtain a first approximation to the power, we evaluate $h(\mu_1)$ from (24) or (25) and enter Table 5. For the second approximation we have to evaluate also $h[\mu_1 \pm \sqrt{(3\mu_2)}]$ and as before obtain the weighted sum of the corresponding P's.

The converse problem of determining the sample sizes for a given power can be solved only if the value of θ is given. For example, we may ask, 'What sample sizes should we take, one being double the other if we want a 90 % chance of detecting a significant difference when p_1 and p_2 have certain specified values?'

First we calculate $k(\theta)$. From the given power we obtain h from Table (5). Then $N = k^2(\theta)/h^2$ will give N to the first approximation. Since θ is given, m and n are obtained.

12. Summary

The power function of the test of significance for the 2×2 table has been considered and an approximate method of deriving it has been developed. The usefulness of the idea of power in fixing in advance the sizes of samples is indicated.

Tables have been provided for determining the power for a specified alternative, the samples sizes being given and conversely.

I wish to express my grateful thanks to Prof. E. S. Pearson and Dr H. O. Hartley for their guidance in the study of this problem.

REFERENCES

Barnard, G. A. (1947). *Biometrika*, **34**, 123.
Beard, R. E. (1947). *J. Inst. Actu.* **73**, 356.
Neyman, J. & Pearson, E. S. (1933). *Philos. Trans.* A, **231**, 289.
Neyman, J. & Tokarska (1936). *J. Amer. Statist. Ass.* **31**, 318.
Pearson, E. S. (1947). *Biometrika*, **34**, 139.
Tang, P. C. (1938). *Statist. Res. Mem.* **2**, 126.

Table 4. *Values of* $k = \frac{\sqrt{2}}{4}\sqrt{\{(p_1+p_2)(2-p_1-p_2)\}}\,\log_e \frac{p_1(1-p_2)}{p_2(1-p_1)}$

[For $p_1 < p_2$, enter with $1-p_1$ and $1-p_2$]

Values of p_2 \ Values of p_1	0	0·05	0·10	0·15	0·20	0·25	0·30	0·35	0·40	0·45	0·50	0·55	0·60	0·65	0·70	0·75	0·80	0·85	0·90	0·95	1·00
1·00																					
0·95																				0	
0·90																			0	0·139	
0·85																		0	0·107	0·257	
0·80																	0	0·094	0·205	0·359	
0·75																0	0·085	0·180	0·295	0·466	
0·70															0	0·079	0·165	0·262	0·382	0·563	
0·65														0	0·076	0·155	0·242	0·342	0·466	0·658	
0·60													0	0·073	0·149	0·230	0·318	0·420	0·549	0·750	
0·55												0	0·072	0·145	0·221	0·303	0·393	0·497	0·630	0·840	
0·50											0	0·071	0·143	0·216	0·294	0·376	0·468	0·574	0·712	0·930	
0·45										0	0·071	0·142	0·214	0·288	0·366	0·450	0·543	0·653	0·794	1·019	
0·40									0	0·072	0·143	0·214	0·287	0·362	0·441	0·526	0·621	0·733	0·878	1·109	
0·35								0	0·073	0·145	0·216	0·288	0·362	0·438	0·518	0·604	0·714	0·815	0·964	1·202	
0·30							0	0·076	0·149	0·221	0·294	0·366	0·441	0·518	0·599	0·687	0·786	0·903	1·055	1·298	
0·25						0	0·079	0·155	0·230	0·303	0·376	0·450	0·526	0·604	0·687	0·777	0·877	0·997	1·152	1·401	
0·20					0	0·085	0·165	0·242	0·318	0·393	0·468	0·543	0·621	0·714	0·786	0·877	0·980	1·102	1·261	1·514	
0·15				0	0·094	0·180	0·262	0·342	0·420	0·497	0·574	0·653	0·733	0·815	0·903	0·997	1·102	1·227	1·388	1·646	
0·10			0	0·107	0·205	0·295	0·382	0·466	0·549	0·630	0·712	0·794	0·878	0·964	1·055	1·152	1·261	1·388	1·554	1·816	
0·05		0	0·139	0·257	0·359	0·466	0·563	0·658	0·750	0·840	0·930	1·019	1·109	1·202	1·298	1·401	1·514	1·646	1·816	2·082	
0																					

Table 5. *Relating Power and k, h, n and α*

h \ α	*P* of equation (19) or the approximate power				Values of $k = h/\sqrt{n}$, according to h and n															n / h
	0·10	0·05	0·02	0·01	10	15	20	25	30	35	40	45	50	60	70	80	90	100	150	
0·1	0·1017	0·0511	0·0206	0·0104*	0·032	0·026	0·022	0·020	0·018	0·017	0·016	0·015	0·014	0·013	0·012	0·011	0·011	0·010	0·008	0·1
0·2	0·1068	0·0546	0·0225	0·0115	0·063	0·052	0·045	0·040	0·037	0·034	0·032	0·030	0·028	0·026	0·024	0·022	0·021	0·020	0·016	0·2
0·3	0·1152	0·0604	0·0257	0·0134	0·095	0·077	0·067	0·060	0·055	0·051	0·047	0·045	0·042	0·039	0·036	0·034	0·032	0·030	0·024	0·3
0·4	0·1270	0·0685	0·0302	0·0162	0·126	0·103	0·089	0·080	0·073	0·068	0·063	0·060	0·057	0·052	0·048	0·045	0·042	0·040	0·033	0·4
0·5	0·1421	0·0791	0·0363*	0·0200	0·158	0·129	0·112	0·100	0·091	0·085	0·079	0·075	0·071	0·065	0·060	0·056	0·053	0·050	0·041	0·5
0·6	0·1604	0·0922	0·0439	0·0248	0·190	0·155	0·134	0·120	0·110	0·101	0·095	0·089	0·085	0·077	0·072	0·067	0·063	0·060	0·049	0·6
0·7	0·1819	0·1077	0·0532	0·0309	0·221	0·181	0·157	0·140	0·128	0·118	0·111	0·104	0·099	0·090	0·084	0·078	0·074	0·070	0·057	0·7
0·8	0·2063	0·1259	0·0643	0·0382	0·253	0·207	0·179	0·160	0·146	0·135	0·126	0·119	0·113	0·103	0·096	0·089	0·084	0·080	0·065	0·8
0·9	0·2336	0·1467	0·0775	0·0471	0·285	0·232	0·201	0·180	0·164	0·152	0·142	0·134	0·127	0·116	0·108	0·101	0·095	0·090	0·073	0·9
1·0	0·2636	0·1701	0·0928	0·0577	0·316	0·258	0·224	0·200	0·183	0·169	0·158	0·149	0·141	0·129	0·120	0·112	0·105	0·100	0·082	1·0
1·1	0·2960*	0·1960*	0·1103	0·0701	0·348	0·284	0·246	0·220	0·201	0·186	0·174	0·164	0·156	0·142	0·131	0·123	0·116	0·110	0·090	1·1
1·2	0·3304	0·2244	0·1302	0·0845†	0·379	0·310	0·268	0·240	0·219	0·203	0·190	0·179	0·170	0·155	0·143	0·134	0·126	0·120	0·098	1·2
1·3	0·3667	0·2552	0·1525†	0·1011	0·411	0·336	0·291	0·260	0·237	0·220	0·206	0·194	0·184	0·168	0·155	0·145	0·137	0·130	0·106	1·3
1·4	0·4044	0·2881	0·1772	0·1199‡	0·443	0·361	0·313	0·280	0·256	0·237	0·221	0·209	0·198	0·181	0·167	0·157	0·148	0·140	0·114	1·4
1·5	0·4432	0·3230	0·2044	0·1410	0·474	0·387	0·335	0·300	0·274	0·254	0·237	0·224	0·212	0·194	0·179	0·168	0·158	0·150	0·122	1·5
1·6	0·4827	0·3596	0·2339	0·1646	0·506	0·413	0·358	0·320	0·292	0·270	0·253	0·239	0·226	0·207	0·191	0·179	0·169	0·160	0·131	1·6
1·7	0·5224	0·3976	0·2656‡	0·1906	0·538	0·439	0·380	0·340	0·310	0·287	0·269	0·253	0·240	0·219	0·203	0·190	0·179	0·170	0·139	1·7
1·8	0·5619	0·4365†	0·2993	0·2189	0·569	0·465	0·402	0·360	0·329	0·304	0·285	0·268	0·255	0·232	0·215	0·201	0·190	0·180	0·147	1·8
1·9	0·6009	0·4761	0·3349	0·2496	0·601	0·491	0·425	0·380	0·347	0·321	0·300	0·283	0·269	0·245	0·227	0·212	0·200	0·190	0·155	1·9
2·0	0·6389	0·5160	0·3721	0·2824	0·632	0·516	0·447	0·400	0·365	0·338	0·316	0·298	0·283	0·258	0·239	0·224	0·211	0·200	0·163	2·0
2·1	0·6756†	0·5557	0·4105	0·3171	0·664	0·542	0·470	0·420	0·383	0·355	0·332	0·313	0·297	0·271	0·251	0·235	0·221	0·210	0·171	2·1
2·2	0·7107	0·5949	0·4497	0·3535	0·696	0·568	0·492	0·440	0·402	0·372	0·348	0·328	0·311	0·284	0·263	0·246	0·232	0·220	0·180	2·2
2·3	0·7439	0·6331	0·4895	0·3913	0·727	0·594	0·514	0·460	0·420	0·389	0·364	0·343	0·325	0·297	0·275	0·257	0·242	0·230	0·188	2·3
2·4	0·7749	0·6701‡	0·5294	0·4302	0·759	0·620	0·537	0·480	0·438	0·406	0·379	0·358	0·339	0·310	0·287	0·268	0·253	0·240	0·196	2·4
2·5	0·8038	0·7054	0·5689	0·4698	0·791	0·645	0·559	0·500	0·456	0·423	0·395	0·373	0·354	0·323	0·299	0·280	0·264	0·250	0·204	2·5
2·6	0·8303‡	0·7389	0·6078	0·5096	0·822	0·671	0·581	0·520	0·475	0·439	0·411	0·388	0·368	0·336	0·311	0·291	0·274	0·260	0·212	2·6
2·7	0·8543	0·7704	0·6457	0·5494	0·854	0·697	0·604	0·540	0·493	0·456	0·427	0·402	0·382	0·349	0·323	0·302	0·285	0·270	0·220	2·7
2·8	0·8760	0·7996	0·6821	0·5887	0·885	0·723	0·626	0·560	0·511	0·473	0·443	0·417	0·396	0·361	0·335	0·313	0·295	0·280	0·229	2·8
2·9	0·8953	0·8264	0·7169	0·6271	0·917	0·749	0·648	0·580	0·529	0·490	0·459	0·432	0·410	0·374	0·347	0·324	0·306	0·290	0·237	2·9
3·0	0·9123	0·8508	0·7497	0·6643	0·949	0·775	0·671	0·600	0·548	0·507	0·474	0·447	0·424	0·387	0·359	0·335	0·316	0·300	0·245	3·0
3·2	0·9400	0·8925	0·8068	0·7337	1·012	0·826	0·716	0·640	0·584	0·541	0·506	0·477	0·453	0·413	0·382	0·358	0·327	0·320	0·261	3·2
3·4	0·9604	0·9251	0·8585	0·7951	1·075	0·878	0·760	0·680	0·621	0·575	0·538	0·507	0·481	0·439	0·406	0·380	0·337	0·340	0·278	3·4
3·6	0·9747	0·9495	0·8986	0·8471	1·138	0·930	0·805	0·720	0·657	0·609	0·569	0·537	0·509	0·465	0·430	0·402	0·348	0·360	0·294	3·6
3·8	0·9844	0·9671	0·9297	0·8896	1·202	0·981	0·850	0·760	0·694	0·642	0·601	0·566	0·537	0·491	0·454	0·425	0·358	0·380	0·301	3·8
4·0	0·9907	0·9793	0·9529	0·9228	1·265	1·033	0·894	0·800	0·730	0·676	0·632	0·596	0·566	0·516	0·478	0·447	0·369	0·400	0·327	4·0
4·2	0·9947	0·9875	0·9695	0·9478	1·328	1·084	0·939	0·840	0·767	0·710	0·664	0·626	0·594	0·542	0·502	0·470	0·379	0·420	0·343	4·2
4·4	0·9971	0·9927	0·9809	0·9659	1·391	1·136	0·984	0·880	0·803	0·744	0·696	0·656	0·622	0·568	0·526	0·492	0·390	0·440	0·359	4·4
4·6	0·9984	0·9959	0·9885	0·9785	1·455	1·188	1·029	0·920	0·840	0·778	0·727	0·686	0·651	0·594	0·550	0·514	0·401	0·460	0·376	4·6
4·8	0·9992	0·9977	0·9933	0·9869	1·518	1·239	1·073	0·960	0·876	0·811	0·759	0·716	0·679	0·620	0·574	0·537	0·411	0·480	0·392	4·8
5·0	0·9996	0·9988	0·9962	0·9923	1·581	1·291	1·118	1·000	0·913	0·845	0·791	0·745	0·707	0·645	0·598	0·559	0·422	0·500	0·408	5·0

THE ANALYSIS OF CONTINGENCY TABLES WITH GROUPINGS BASED ON QUANTITATIVE CHARACTERS

By F. YATES

A $p \times q$ contingency table can be tested for independence by a χ^2 test with $(p-1)(q-1)$ degrees of freedom. This is an over-all test which covers all forms of departure from proportionality, and is consequently correspondingly insensitive to departures of a specified type.

If the nature of the data is such that departures of a particular type are to be expected, then a test of significance appropriate to departures of this type will be justified.

The present paper deals with the case in which one or both groupings are based on characters which are either directly quantitative or are in the form of gradings which can be regarded as having an underlying quantitative basis. The actual data which gave rise to the investigation are shown in Table 1. They were obtained in the course of a pilot inquiry into the conditions in which school children do their homework, carried out by the Department of Social Science, University of Liverpool, and I am indebted to Mr D. Chapman for permission to reproduce them here.

Table 1. *Relation (in terms of numbers of children and percentages) between conditions under which homework was carried out, and the teacher's rating of the quality of that homework. (Each scale is graded, A being the highest rating)*

Teacher's rating	Homework conditions					
	A	*B*	*C*	*D*	*E*	Total
A	141 (46%)	67 (46%)	114 (39%)	79 (44%)	39 (43%)	440 (43%)
B	131 (42%)	66 (45%)	143 (48%)	72 (40%)	35 (39%)	447 (44%)
C	36 (12%)	14 (9%)	38 (13%)	28 (16%)	16 (18%)	132 (13%)
Total	308 (100%)	147 (100%)	295 (100%)	179 (100%)	90 (100%)	1019 (100%)

It is clear from the percentages that the effect of homework conditions on the quality of the preparation, as judged by the teacher's rating, is small. On the other hand, there is some slight trend, and the question therefore arises whether this trend has any significance, or, more generally, what is its estimated magnitude, and what are the errors of this estimate.

A χ^2 test of the whole table gives a value of χ^2 equal to 9·16 (8 degrees of freedom), but such a test, as pointed out above, embraces all types of deviation from proportionality.

In material of this kind a rough test, which isolates the major part of any quantitative association between the two variates, can be made by reduction of the table to a 2×2 table by the grouping of appropriate parts and rejection of others. With the above data we might reasonably group homework conditions A and B, and D and E, rejecting condition C, and also reject the teacher's central rating B. This will give the values of Table 2.

This Table gives a value of χ_c^2, i.e. χ^2 corrected for continuity, equal to 3·03 (1 degree of freedom), corresponding to a probability (for one tail of the χ_c distribution) of 0·041. (The use of a single tail is appropriate to testing whether there is evidence that improved conditions of homework improve the quality of preparation, ruling out the opposite contingency.) The test therefore indicates that there is significant evidence (at the 5 % level) of some improvement.

The above rough test, however, is open to several objections. In the first place, since the grouping is arbitrary, there is always a possibility that the statistician will allow himself to be influenced by the data in his choice of grouping, a grouping which gives a high degree of association being chosen. If this occurs, the test of significance is clearly vitiated. Secondly, different workers may in good faith choose different groupings of the same data, and this may lead to arguments of a type that are likely to discredit the science of statistics. Thirdly, the choice of grouping which is most appropriate in any given case depends on the marginal totals of the numbers of observations, and simple rules for the choice of grouping are not easy to devise. Fourthly, the test provides no estimate of the magnitude of the effects of the association.

Table 2. *Condensation of part of Table* 1

Teacher's rating	Homework conditions		
	$A+B$	$D+E$	Total
A C	208 (81 %) 50 (19 %)	118 (73 %) 44 (27 %)	326 (78 %) 94 (22 %)
Total	258 (100 %)	162 (100 %)	420 (100 %)

The test based on regression concepts, developed below, eliminates the element of choice, requires no elaborate computation, and also provides estimates of the magnitude of the effect of each variate on the other.

The following notation will be adopted. Letters without dashes will be taken to indicate the results of operations on the rows, and with dashes the results of the same operations on the columns. The r row totals will be denoted by $N_1, N_2, \ldots$, the r' column totals by $N'_1, N'_2, \ldots$, and the grand total by T. We shall also require an extended summation notation, in which the x's may denote any set of r quantities, as follows,

$$\begin{aligned} S_0(x) &= x_1 + x_2 + \ldots + x_r, \\ S_1(x) &= -\tfrac{1}{2}(r-1)\,x_1 - \tfrac{1}{2}(r-3)\,x_2 - \ldots + \tfrac{1}{2}(r-1)\,x_r, \\ S_2(x) &= \tfrac{1}{4}(r-1)^2 x_1 + \tfrac{1}{4}(r-3)^2 x_2 + \ldots + \tfrac{1}{4}(r-1)^2 x_r, \end{aligned}$$

with three similar functions extending to r' terms distinguished by dashes.

For each of the r' columns a mean score is calculated, assigning a score of $-\tfrac{1}{2}(r-1)$ to the first row, $-\tfrac{1}{2}(r-3)$ to the second row, etc. and $+\tfrac{1}{2}(r-1)$ to the last row. If the numbers in the different sub-groups of the first column are $n_1, n_2, \ldots, n_r$, the mean score for the column is

$$u'_1 = \frac{1}{N'_1} S_1(n),$$

and the total score is

$$U'_1 = S_1(n).$$

If the numbers $n_1, n_2, \ldots, n_r$ are regarded as a sample from a multinomial distribution with probabilities of, say, $p_1, p_2, \ldots, p_r$, the variance of n_1 will be $N'_1 p_1 q_1$, and the covariance of n_1 and n_2 will be $-N'_1 p_1 p_2$, etc. The variance of u'_1 can then be simply calculated, and will be found to be

$$V(u'_1) = \frac{1}{N'_1}[S_2(p) - \{S_1(p)\}^2].$$

The regression of the mean scores $u'_1, u'_2, \ldots, u'_{r'}$ for the r' columns on column number can now be calculated, weighting according to the variances of the u''s, provided we can make some assumption as to the values of the p's.

In order to obtain a test of significance we may start with the hypothesis that the sets of p's are identical for all columns, i.e. that p_1 is the same from column to column, etc. Then the variances of the u''s are inversely proportional to the numbers in the columns, i.e. the N''s, and in calculating the regression we may therefore take weights equal to the N''s. Estimates of the values of p's may be derived from the row totals, so that $p_1 = N_1/T$, etc., when we shall have

$$V(u'_1) = \frac{1}{N'_1 T^2}[TS_2(N) - \{S_1(N)\}^2],$$

etc.

If the regression equation is taken in the form

$$u'_x = m' + b'\{x - \tfrac{1}{2}(r+1)\},$$

the equations of estimation for m' and b' are:

$$m'S'_0(N') + b'S'_1(N') = S'_0(U'), \quad m'S'_1(N') + b'S'_2(N') = S'_1(U').$$

We have

$$S_0(N) = S'_0(N') = T, \quad S'_0(U') = S_1(N),$$

$$S_0(U) = S'_1(N'), \qquad S_1(U) = S'_1(U').$$

If we put

$$A = TS_2(N) - \{S_1(N)\}^2,$$

$$A' = TS'_2(N') - \{S'_1(N')\}^2,$$

$$B = TS_1(U) - S_1(N)S'_1(N'),$$

the solution of the above equations gives $b' = B/A'$, with the corresponding regression on row scores $b = B/A$.

The variance per unit weight is A/T^2, and consequently

$$V(b') = \frac{T}{A'}\frac{A}{T^2} = \frac{A}{A'T}.$$

Similarly, $V(b) = A'/AT$. The appropriate test of significance is therefore given by

$$\chi^2 = \frac{b^2}{V(b)} = \frac{b'^2}{V(b')} = \frac{B^2 T}{AA'}$$

with one degree of freedom.

The above test is unaffected by interchange of rows and columns, as should be the case. It can easily be verified that the test reduces to the ordinary χ^2 test of a 2×2 contingency table (apart from the correction for continuity) if the values in all but two of the rows and two of the columns are put equal to zero. The correction for continuity is not here of importance in view of the large number of possible alternatives with given marginal totals.

The test has been arrived at by assuming that the column totals only are fixed, but with the additional approximation involved in assuming that the p's for each column are given by the marginal totals for the rows. This approach has the advantage of indicating the

appropriate criterion B^2T/AA' for the type of association which it is desired to test. Once the criterion is determined, however, it can be shown, following Fisher (1922, 1925), that this criterion, which is linear in the deviations from expectation, and orthogonal with the linear functions representing the row and column totals, will in large samples be distributed as χ^2 with 1 degree of freedom when both sets of marginal totals are held fixed.

The values for the regression coefficients b and b' give estimates of the change in mean score with unit change of row and column respectively. It should be noted that the variances of b and b' given above are based on the assumption that there is no association between the two variates and become progressively more inaccurate as the degree of association increases. In general they will be over-estimates of the true variances. Nothing more accurate is likely to be required, however, except when there is very marked association, in which case the assumption of linear regression in the mean scores over the whole range is unlikely to provide an adequate mathematical description of the association.

If only one of the classifications, say the rows, is quantitative, a test for the homogeneity of the mean scores for the columns may be derived from the variances of these mean scores. The quantity

$$\begin{aligned} Q &= S'_0\left(\frac{u'^2}{V(u')}\right) - \frac{[S'_0\{u'/V(u')\}]^2}{S'_0\{1/V(u')\}} \\ &= \frac{T^2}{A}\left[S'_0\left(\frac{U'^2}{N'}\right) - \frac{\{S'_0(U')\}^2}{T}\right] \\ &= \frac{T^2}{A}\,[S'_0(u'U') - \bar{u}'S'_0(U')], \end{aligned}$$

where $\bar{u}'$ is the mean score for all the observations, will be distributed as χ^2 with $r'-1$ degrees of freedom. It is easily verified that this test reduces to the ordinary χ^2 test for a $2\times r'$ table when there are only two rows.

It may be noted that any system of scoring may be assigned to each of the classifications—there is no need to adopt a system with equal intervals between each class if the nature of the classification is such that scoring with unequal intervals is more appropriate. If, for example, in the data of Table 1 the teachers had given their opinion that the difference between their classes A and B in quality of preparation was only half that between B and C scores of $+1, 0$, and -2 might have been adopted. In the absence of any such indications, however, the appropriate procedure will be to assume that graded classifications are intended to represent equal intervals on some scale, unless the data themselves are used to determine the optimal system of scoring, by some procedure analogous to that given by Fisher (1946), para. 49·2, or by reference to some assumed distribution of the scores in the population. Examples of the latter procedure are described by Pearson & Moul (1925), and also in K. Pearson's *Tables for Statisticians and Biometricians*, Part II, pp. xxiii–xxvi (1931). In this connexion see also E. S. Pearson (1923). Such procedures, however, are only likely to be worth while in exceptional cases, and with very extensive data.

The computational procedure when the scores are assumed is very simple, and is illustrated in Table 3 for a 4×5 table with quantitative classifications. The scores used for the rows have been multiplied by 2 for convenience in computation. The value of b obtained will therefore represent half the change to be expected in the mean score from row to row.

The total scores U and U' in Table 3 are calculated by summing the products of the scores and the numbers in the corresponding sub-classes. Checks are provided by carrying out the same operations on the totals. $S_2(N)$ and $S_2(N')$ are obtained by multiplying the totals

$N_1, N_2, \ldots$, and $N'_1, N'_2, \ldots$, by the squares of the scores. These must be checked. $S_1(U)$ is obtained in two ways by summing the products of the U's and the U''s with their corresponding scores.

The quantity A is then obtained from the last three values in the total column, A' from the last three values in the total line, and B from the cross product of the 2×2 table formed by the total and total score rows and columns.

Table 3. *Computational procedure*

		(Score)^2	4	1	0	1	4			
		Score	-2	-1	0	$+1$	$+2$			
(Score)^2	Score		*A*	*B*	*C*	*D*	*E*	Total	Total score	
9	-3	*A*	—	—	—	—	—	N_1	U_1	
1	-1	*B*	—	—	—	—	—	N_2	U_2	
1	$+1$	*C*	—	—	—	—	—	N_3	U_3	
9	$+3$	*D*	—	—	—	—	—	N_4	U_4	
		Total	N'_1	N'_2	N'_3	N'_4	N'_5	T	$S_0(U) = S'_1(N')$	$S_2(N') \leftarrow A'$
		Total score	U'_1	U'_2	U'_3	U'_4	U'_5	$S'_0(U') = S_1(N)$	$S_1(U) = S'_1(U')$ $\nwarrow B$	
								$S_2(N)$ $\uparrow$ A		

Table 4. *Analysis of the data of Table* 1

			Homework conditions								
		(Score)^2	4	1	0	1	4				
		Score	$+2$	$+1$	0	-1	-2				
(Score)^2	Score	Teacher's rating	*A*	*B*	*C*	*D*	*E*	Total	Total score	Total (score)^2	Mean score
1	$+1$	*A*	141	67	114	79	39	440	$+192$		$+0{\cdot}44$
0	0	*B*	131	66	143	72	35	447	$+186$		$+0{\cdot}42$
1	-1	*C*	36	14	38	28	16	132	$+26$		$+0{\cdot}20$
		Total	308	147	295	179	90	1019	$+404$	1918	$+0{\cdot}40$
		Total score	$+105$	$+53$	$+76$	$+51$	$+23$	$+308$	$+166$		
		Total (score)^2						572			
		Mean score	$+0{\cdot}34$	$+0{\cdot}36$	$+0{\cdot}26$	$+0{\cdot}28$	$+0{\cdot}26$	$+0{\cdot}30$			

The analysis of the data of Table 1 is given in Table 4, which also shows the mean scores (not required in the analysis). From the values of this table we have:

$$A = 1019 \times 572 - 308^2 = 488{,}004,$$

$$A' = 1019 \times 1918 - 404^2 = 1{,}791{,}226,$$

$$B = 1019 \times 166 - 308 \times 404 = 44{,}722,$$

$$b = B/A = 44{,}722/488{,}004 = 0{\cdot}09164,$$

$$b' = B/A' = 44{,}722/1{,}791{,}226 = 0{\cdot}02497,$$

$$\text{S.E. of } b = \sqrt{\frac{A'}{AT}} = \sqrt{\frac{1{,}791{,}226}{488{,}004 \times 1019}} = \pm 0{\cdot}0600,$$

$$\text{S.E. of } b' = \sqrt{\frac{A}{A'T}} = \sqrt{\frac{488{,}004}{1{,}791{,}226 \times 1019}} = \pm 0{\cdot}0163,$$

$$\chi^2 = \frac{B^2T}{AA'} = \frac{44{,}722^2 \times 1019}{488{,}004 \times 1{,}791{,}226} = 2{\cdot}332.$$

Reference to a table of the normal integral indicates that $P = 0{\cdot}062$ (one tail), and there is therefore a probability about 1 in 16 of obtaining by chance as great an apparent improvement as is indicated by the data, if homework conditions have in fact no influence on the quality of the homework as shown by the teacher's rating. More important, in data of this kind, the analysis indicates that the amount of improvement, if any, is likely to be small. In terms of the improvement to be expected in changing conditions from the worst (E) to the best (A) the estimated improvement, $4b'$, is $0{\cdot}100$, the limits given by once and twice the standard error being $+0{\cdot}03$ to $+0{\cdot}16$ and $-0{\cdot}03$ to $+0{\cdot}23$. We may therefore state that the true value of this improvement is not likely to be substantially greater than $+0{\cdot}16$ and is almost certainly not greater than $+0{\cdot}23$, nor is it likely to be substantially less than $+0{\cdot}03$, and is almost certainly not less than $-0{\cdot}03$.

It will be noted that the exact test in this instance gives a lower degree of significance than the rough test based on reduction to a 2×2 table. This, of course, does not imply that the exact test is less sensitive. With any given set of data the verdicts of different tests of significance may differ considerably owing to chance causes. This does not preclude the use of rough tests when considerations of speed and computational labour are paramount, but once the exact test has been evaluated the verdict of the simpler but less appropriate test must be set aside. The decision as to the test to be used should also be made without reference to the actual data—it is, of course, inadmissible to make a rough test, accept its verdict if significant, and proceed to an exact test if it fails to indicate significance.

As an example of the computations when one classification only is quantitative we may test the deviations between the mean scores for different homework conditions. This test would be required if the homework conditions were qualitative categories and not an ordered series. In this case the value of χ^2 for 4 degrees of freedom will be given by

$$\begin{aligned} Q &= \frac{1019^2}{488{,}004}(105 \times 0{\cdot}34091 + \ldots - 308 \times 0{\cdot}302257) \\ &= 3{\cdot}825, \end{aligned}$$

the mean scores being taken to 5 and 6 decimal places to ensure the necessary accuracy. It is clear that apart from the linear trend there are no significant variations in mean score.

REFERENCES

FISHER, R. A. (1922). *J. R. Statist. Soc.* **85**, 87.
FISHER, R. A. (1925). *Metron*, **5**, 3.
FISHER, R. A. (1946). *Statistical Methods for Research Workers*, 10th ed. Edinburgh: Oliver and Boyd.
PEARSON, E. S. (1923). *Biometrika*, **7**, 248.
PEARSON, K. (1931). *Tables for Statisticians and Biometricians*, Part II. London: Biometrika Office.
PEARSON, K. & MOUL, M. (1925). *Ann. Eugen., Lond.*, **1**, 1.

THE PROBABILITY INTEGRAL TRANSFORMATION WHEN PARAMETERS ARE ESTIMATED FROM THE SAMPLE

By F. N. DAVID and N. L. JOHNSON

1. The probability integral transformation for testing goodness of fit and combining tests of significance was introduced by R. A. Fisher in 1932. Fisher's objective was the significance of combined independent tests of significance, but his method also proved applicable to a certain limited range of tests for goodness of fit as can be seen in K. Pearson (1933), J. Neyman (1937) and E. S. Pearson (1938). The transformation may be summarized briefly in the following way. Assume that there is a continuous random variable x whose elementary probability law is $p(x)$, whence obviously

$$\int_{-\infty}^{+\infty} p(x)\,dx = 1.$$

Consider a new random variable, y, connected with x by the relation

$$y = \int_{-\infty}^{x} p(x)\,dx.$$

y is a monotonic non-decreasing function of x and $0 \leqslant y \leqslant 1$. Further, if $p(y)$ is the elementary probability law of y, then

$$p(y) = p(x)\frac{dx}{dy} = 1.$$

Hence in the interval [0; 1] all values of y are equally likely, or in common parlance, y is rectangularly distributed in the interval [0; 1], no matter what the elementary probability law of x. If therefore we have n independent random variables $x_j (j = 1, 2, \ldots, n)$ following a known continuous probability law which is completely specified by H_0, the hypothesis tested, then by means of the transformation

$$y_j = \int_{-\infty}^{x_j} p(x_j \mid H_0)\,dx_j,$$

the x's can be transformed into n independent random variables y which are rectangularly distributed.

2. The transformation which we have just summarized is useful statistically in that tests based on a rectangular population can be made applicable to any variable of which the elementary probability law is known. However, because the parameters of the elementary probability law must be specified, it is clear that the range of application of any tests based on this transformation will be very restricted, for cases are rare in statistical practice when H_0 is completely specified. It seemed interesting to us to investigate the effect on the transformation of calculating estimates of the parameters from the data provided by the sample. For example, if the mean of the probability law is estimated from a sample of n quantities $X_1, X_2, \ldots, X_n$ each of which is one observed value of n random variables $x_1, x_2, \ldots, x_n$, the y's obtained by the probability integral transformation will no longer be independent, neither will they be rectangularly distributed. We are able to show that the generality of the transformation in the case when the parameters are completely specified is lost as soon as we begin replacing unknown parameters by the sample estimates, and, as is intuitively obvious, the form of the probability law of y depends on the functional form of the common probability law of the x's.

3. We begin by stating the problem in a formal mathematical way and indicate the method whereby a general solution is reached. Assume that $p(x)$ is a single valued continuous function of the form

$$p(x) = f(x \mid \theta_1, \theta_2, \ldots, \theta_s),$$

where, in the usual way, $\theta_1, \theta_2, \ldots, \theta_s$, are parameters descriptive of the population, all of which may or may not be specified. It may be assumed for generality that none are specified and that in place of the unknown parameters, θ, it is necessary to substitute functions of the sample values, say,

$$F_1(x_1, x_2, \ldots, x_n), \quad F_2(x_1, x_2, \ldots, x_n), \quad \ldots, \quad F_s(x_1, x_2, \ldots, x_n),$$

where $x_1, x_2, \ldots, x_n$ are the random variables of which the n observations which form the sample are the observed values. Thus we require to find the distribution of the variables

$$y_i = \int_{-\infty}^{x_i} f(t \mid F_1, F_2, \ldots, F_s)\, dt \quad \text{for} \quad i = 1, 2, \ldots, n.$$

We have immediately that

$$\frac{\partial y_i}{\partial x_j} = \int_{-\infty}^{x_i} \sum_{r=1}^{s} \frac{\partial f}{\partial F_r} \frac{\partial F_r}{\partial x_j} dt = \sum_{r=1}^{s} \frac{\partial F_r}{\partial x_j} \int_{-\infty}^{x_i} \frac{\partial f}{\partial F_r} dt \quad \text{for} \quad i \neq j,$$

and

$$\frac{\partial y_i}{\partial x_i} = f(x_i \mid F_1, F_2, \ldots, F_s) + \sum_{r=1}^{s} \frac{\partial F_r}{\partial x_i} \int_{-\infty}^{x_i} \frac{\partial f}{\partial F_r} dt \quad \text{for} \quad i = j.$$

In matrix notation we may write this

$$\left[\frac{\partial(y_1, \ldots, y_n)}{\partial(x_1, \ldots, x_n)}\right] = \Delta_{\mathrm{I}} + \left[\frac{\partial F_j}{\partial x_k}\right] \left[\int_{-\infty}^{x_j} \frac{\partial f}{\partial F_k} dt\right].$$

Δ_{I} is a diagonal matrix with diagonal elements $f(x_i \mid F_1, \ldots, F_s)$. $[\partial F_j/\partial x_k]$ is an $n \times s$ matrix, $\partial F_j/\partial x_k$ being the element in the kth row and the jth column. $\left[\int_{-\infty}^{x_j} \frac{\partial f}{\partial F_k} dt\right]$ is an $s \times n$ matrix, $\int_{-\infty}^{x_j} \frac{\partial f}{\partial F_k} dt$ being the element in the kth row and the jth column. The rank of the matrix $\left[\frac{\partial(y_1, \ldots, y_n)}{\partial(x_1, \ldots, x_n)}\right]$ is not immediately obvious. In general it may be noted that it will be at least $n-s$, and that it will be less than n. A study of particular cases leads us to believe that where the s sample estimates are algebraic functions each of the other, as for example, the sample moment coefficients, then there will be s independent relationships between the y's, and the rank of the matrix will be $n-s$. Where the sample estimates are not functions of one another, as for example in the case of the median and the standard deviation, the matrix rank will be between $n-s$ and n. We have not been able to prove this in general but it should not be impossible to do so.

We shall assume that there are s independent relationships between the variables $y_1, y_2, \ldots, y_n$. Under this last assumption we have

$$\frac{\partial(y_1, \ldots, y_{n-s}, F_1, F_2, \ldots, F_s)}{\partial(x_1, \ldots, x_n)} = \frac{\partial(F_1, \ldots, F_s)}{\partial(x_{n-s+1}, \ldots, x_n)} \prod_{i=1}^{n-s} f(x_i \mid F_1, F_2, \ldots, F_s)$$

(provided partial differentiation is, in fact, possible), whence, since

$$p(x_1, x_2, \ldots, x_n \mid \theta_1, \theta_2, \ldots, \theta_s) = \prod_{i=1}^{n} f(x_i \mid \theta_1, \theta_2, \ldots, \theta_s),$$

the joint probability law of $y_1, y_2, \ldots, y_{n-s}, F_1, \ldots, F_s$, may be written

$$p(y_1, y_2, \ldots, y_{n-s}, F_1, F_2, \ldots, F_s \mid \theta_1, \theta_2, \ldots, \theta_s) \quad \frac{\prod_{i=1}^{n} f(x_i \mid \theta_1, \theta_2, \ldots, \theta_s)}{\prod_{i=1}^{n-s} f(x_i \mid F_1, F_2, \ldots, F_s)} \quad \frac{\partial(F_1, \ldots, F_s)}{\partial(x_{n-s+1}, \ldots, x_n)}^{-1}$$

Alternative expressions for the joint-probability law may be obtained by using the relationship

$$\prod_{i=1}^{n-s} f(x_i \mid \theta_1, \ldots, \theta_s) = p(x_1, \ldots, x_{n-s} \mid \theta_1, \ldots, \theta_s).$$

Substituting in the right-hand side of the joint law we have

$$\begin{aligned} p(y_1, y_2, \ldots, y_{n-s}, F_1, F_2, \ldots, F_s \mid \theta_1, \theta_2, \ldots, \theta_s) \\ &= \prod_{i=1}^{n-s} \frac{f(x_i \mid \theta_1, \theta_2, \ldots, \theta_s)}{f(x_i \mid F_1, F_2, \ldots, F_s)} \prod_{i=n-s+1}^{n} f(x_i \mid \theta_1, \theta_2, \ldots, \theta_s) \left| \frac{\partial(F_1, \ldots, F_s)}{\partial(x_{n-s+1}, \ldots, x_n)} \right|^{-1} \\ &= \prod_{i=1}^{n-s} \frac{f(x_i \mid \theta_1, \theta_2, \ldots, \theta_s)}{f(x_i \mid F_1, F_2, \ldots, F_s)} p(F_1, F_2, \ldots, F_s \mid x_1, \ldots, x_{n-s+1}, \theta_1, \ldots, \theta_s) \\ &= \frac{p(x_1, \ldots, x_{n-s}, F_1, F_2, \ldots, F_s \mid \theta_1, \ldots, \theta_s)}{\prod^{n-s} f(x_i \mid F_1, F_2, \ldots, F_s)}. \end{aligned}$$

4. In the previous section formal solutions only of the problem have been set down. For any particular case the analysis becomes somewhat complicated. Accordingly, in order to obtain a clear idea of the kinds of distributions arising, we shall first confine ourselves to the discussion of (α) the special case where only location and scale parameters appear in the probability law of the x's and (β) the distribution of single y_i. Under (β) we may note that

$$y_i = \int_{-\infty}^{x_i} f(t \mid F_1, F_2, \ldots, F_s)\, dt = Z_i(x_i, F_1, \ldots, F_s), \quad \text{say.}$$

If the distribution of Z_i is known, the distribution of y_i may be found immediately, and in particular if we can write

$$y_i = Z_i(x_i, F_1, \ldots, F_s) = \int_{-\infty}^{z_i(x_i, F_1, \ldots, F_s)} g(t)\, dt,$$

then

$$p(y_i) = (g(z_i))^{-1} p(z_i).$$

5. It is not uncommon in statistical practice to find probability laws which are completely specified by a single parameter for location and a single parameter for scale. The normal curve is, of course, the classic example. Let ξ be the parameter of location, σ the scale parameter, and write

$$p(x) = f(x \mid \xi, \sigma).$$

If

$$y_i = \int_{-\infty}^{x_i} f(t \mid \xi, \sigma)\, dt = \int_{-\infty}^{(x_i-\xi)/\sigma} f(t \mid 0, 1)\, dt = \int_{-\infty}^{(x_i-\xi)/\sigma} f(t)\, dt, \quad \text{say,}$$

then we may write

$$\frac{x_i - \xi}{\sigma} = \phi(y_i).$$

Either ξ, or σ, or both may be estimated from the observed values of the random variables x. We treat two distinct cases.

Case (i): σ known and ξ estimated

We first suppose that the scaling parameter is known but that it is necessary to estimate a central measure of location. Suppose this to be a function $M(x_1, x_2, \ldots, x_n)$ which may be written for brevity $M(x)$. We have

$$y_i = \int_{-\infty}^{x_i} f(t \mid M(x), \sigma)\, dt = \int_{-\infty}^{[x_i - M(x)]/\sigma} f(t)\, dt. \tag{1}$$

so

$$\frac{x_i - M(x)}{\sigma} = \phi(y_i).$$

It follows that

$$M[\phi(y)] = 0,$$

provided M satisfies the usual conditions for a measure of location.* In this case therefore there is one relation between the y_i's and $\left[\frac{\partial(y_1, \ldots, y_n)}{\partial(x_1, \ldots, x_n)}\right]$ is of rank $n-1$. Using the general formula obtained in § 3, the joint-probability law of $y_1, y_2, \ldots, y_{n-1}, \bar{x}$, is

$$p(y_1, y_2, \ldots, y_{n-1}, \bar{x}) = n \frac{\prod_{i=1}^{n} f(x_i \mid \xi, \sigma)}{\prod_{i=1}^{n-1} f(x_i \mid \bar{x}, \sigma)}$$

The distribution of any individual y_i is simply obtained. For, since

$$y_i = \int_{-c}^{[x_i - M(x)]/\sigma} f(t)\, dt,$$

we have

$$p(y_i) = \left[f\left(\frac{x_i - M(x)}{\sigma}\right)\right]^{-1} p\left(\frac{x_i - M(x)}{\sigma}\right)$$

and if the distribution of $x_i - M(x)$ is known, the distribution of y_i follows immediately.

Case (ii): both ξ and σ estimated

Assume that ξ is estimated as before by $M(x_1, x_2, \ldots, x_n) = M(x)$. Since now σ is also unknown, suppose that it is estimated from the sample values by a measure of dispersion, say

$$D(x_1, x_2, \ldots, x_n) = D(x).$$

We have

$$y_i = \int_{-\infty}^{x_i} f(t \mid M(x), D(x))\, dt,$$

and

$$\frac{x_i - M(x)}{D(x)} = \phi(y_i).$$

Provided $D(x)$ is a function of the quantities $x_i - M(x)$ and satisfies the usual conditions for a measure of dispersion,† the y_i's must now satisfy the two conditions,

$$M[\phi(y)] = 0; \quad D[\phi(y)] = 1.$$

The matrix $\left[\frac{\partial(y_1, \ldots, y_n)}{\partial(x_1, \ldots, x_n)}\right]$ is then of rank $n-2$. The conditions are satisfied, for example, if $M(x) = \bar{x}$ and $D(x) = \left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{\frac{1}{2}}$. By an argument precisely similar to that of case (i) we have that

$$p(y_i) = \left[f\left(\frac{x_i - M(x)}{D(x)}\right)\right]^{-1} p\left(\frac{x_i - M(x)}{D(x)}\right),$$

* $M(x_1 + a, x_2 + a, \ldots, x_n + a) = M(x_1, x_2, \ldots, x_n) + a$; $M(x, x, x, \ldots, x) = x$.

† (i) $D(x_1 + a, x_2 + a, \ldots, x_n + a) = D(x_1, x_2, \ldots, x_n)$;

(ii) $D(x, x, \ldots, x) = 0$;

(iii) $D(kx_1, kx_2, \ldots, kx_n) = |k|\, D(x_1, x_2, \ldots, x_n)$.

where
$$y_i = \int_{-\infty}^{[x_i - M(x)]/D(x)} f(t)\,dt. \tag{2}$$

It is seen that y_i is rectangularly distributed, i.e. $p(y_i) = 1$, if and only if
$$f\left(\frac{x_i - M(x)}{D(x)}\right) = p\left(\frac{x_i - M(x)}{D(x)}\right).$$
This condition is not likely to be satisfied.

For both cases (i) and (ii) it may be noted that $p(y_i)$ is the ratio of two probability laws with a transformation of variables given by (1) and (2) respectively, and that *neither of these two probability laws depends on ξ or on σ.*

6. *Example I.* Let
$$p(x) = \frac{1}{\sqrt{(2\pi)}\,\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\xi}{\sigma}\right)^2\right] = f(x \mid \xi, \sigma).$$
and as in the previous section consider two cases.

For case (i) there are many good statistical reasons for choosing
$$M(x) = \bar{x}$$
for the estimate of ξ for this probability law. In the notation of §4
$$z_i = \frac{x_i - \bar{x}}{\sigma}$$
and
$$p\left(\frac{x_i - \bar{x}}{\sigma}\right) = p(z_i) = \frac{1}{\sqrt{(2\pi)}}\sqrt{\frac{n}{n-1}}\exp\left[-\frac{nz_i^2}{2(n-1)}\right].$$
Applying the results of the preceding section we shall have
$$p(y_i) = \left(\frac{1}{\sqrt{(2\pi)}}\exp\left[-\frac{z_i^2}{2}\right]\right)^{-1}\left(\frac{1}{\sqrt{(2\pi)}}\sqrt{\frac{n}{n-1}}\exp\left[-\frac{nz_i^2}{2(n-1)}\right]\right) = \sqrt{\frac{n}{n-1}}\exp\left[-\frac{z_i^2}{2(n-1)}\right], \tag{3}$$
where $z_i = \phi(y_i)$ is defined by
$$y_i = \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{z_i} e^{-\frac{1}{2}t^2}\,dt.$$
Clearly $p(y_i)$ has a maximum value $\sqrt{n/(n-1)}$ at $z_i = 0$, i.e. when $y = \frac{1}{2}$, and the probability law is symmetrical about this point. $p(y_i)$ is zero at the points $y_i = 0 (z_i = -\infty)$ and $y_i = 1 (z_i = +\infty)$. A graph of the function for three different values of n is given in Fig. 1. In order to compare $p(y_i)$ with the rectangular distribution we may find the points at which the curve crosses it. This will be when $p(y_i) = 1$ or when
$$\frac{z_i^2}{2(n-1)} = -\tfrac{1}{2}\log\left(1 - \frac{1}{n}\right).$$
Expanding the logarithm as a series we have that
$$z_i^2 \simeq 1 - \frac{1}{2n} + \frac{1}{12n^2} \ldots,$$
or, for n moderately large, z_i is nearly equal to ± 1. It follows that $p(y_i) = 1$ when $y_i \simeq 0{\cdot}159$ or $0{\cdot}841$.

In case (ii) for the same $p(x)$, assume
$$M(x) = \bar{x}, \quad D(x) = s = \left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{\frac{1}{2}}.$$

As before, write
$$z_i = \frac{x_i - \bar{x}}{s}$$
then
$$p(z_i) = \frac{\sqrt{n}}{(n-1)\,B\left(\frac{1}{2}, \frac{n-2}{2}\right)}\left(1 - \frac{nz_i^2}{(n-1)^2}\right)^{\frac{1}{2}(n-4)} \quad \text{for} \quad -\frac{n-1}{\sqrt{n}} < z_i < +\frac{n-1}{\sqrt{n}}.$$

It follows that
$$p(y_i) = \left(\frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}z_i^2}\right)^{-1}\left(\frac{\sqrt{n}}{(n-1)\,B\left(\frac{1}{2}, \frac{n-2}{2}\right)}\left(1 - \frac{nz_i^2}{(n-1)^2}\right)^{\frac{1}{2}(n-4)}\right)$$
$$= \frac{\sqrt{(2\pi n)}}{n-1}\,\frac{1}{B(\frac{1}{2}, \frac{1}{2}(n-2))}\left(1 - \frac{nz_i^2}{(n-1)^2}\right)^{\frac{1}{2}(n-4)} e^{\frac{1}{2}z_i^2}, \tag{4}$$

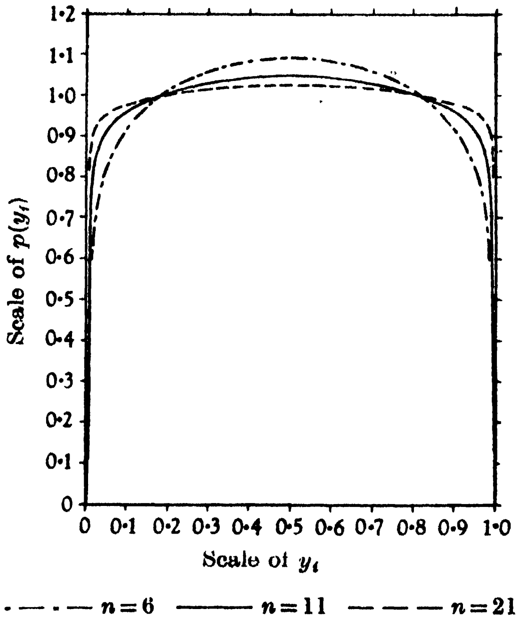


Fig. 1

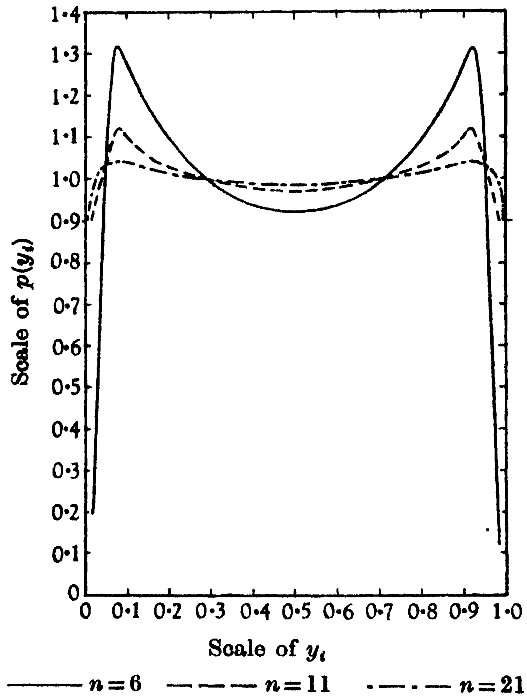


Fig. 2

Fig. 1. The probability integral transformation applied to the normal curve with estimated mean.
Fig. 2. The probability integral transformation applied to the normal curve with estimated mean and standard deviation.

$$\left(\text{Maxima approximately at } y_i = \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{\pm\sqrt{[2+(1/n)]}} e^{-\frac{1}{2}t^2}\,dt.\right)$$

where, as before,
$$y_i = \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{z_i} e^{-\frac{1}{2}t^2}\,dt.$$

A graph of this function, for the same sample sizes considered in case (i), is given in Fig. 2.

7. *Example II*. Let x be distributed as χ^2 with two degrees of freedom, i.e. let
$$p(x) = \frac{1}{\theta}e^{-x/\theta} = f(x \mid \theta) \quad \text{for} \quad x > 0.$$

If we estimate θ by
$$M(x) = \bar{x},$$
then
$$y_i = \frac{1}{\bar{x}}\int_0^{x_i} e^{-t/\bar{x}}\,dt = 1 - e^{-x_i/\bar{x}}.$$
In this case we know, writing
$$u_i = x_i/\bar{x},$$
that
$$p(u_i) = \frac{n-1}{n}\left(1-\frac{u_i}{n}\right)^{n-2} \quad (0 < u_i < n).$$

Following the procedure of the previous sections, we have that
$$p(y_i) = \frac{n-1}{n}\frac{1}{1-y_i}\left[1+\frac{1}{n}\log(1-y_i)\right]^{n-2}$$
for
$$0 < y_i < 1-e^{-n}. \qquad (5)$$

A graph of this function is given in Fig. 3.

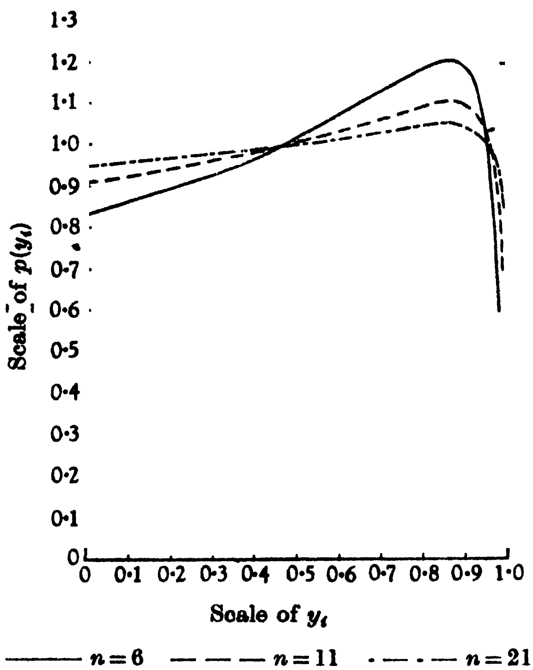


Fig. 3. The probability integral transformation applied to the law
$$p(x) = \frac{1}{\theta}e^{-x/\theta},$$
θ being estimated from the data. (The maximu is at $1-e^{-2} = 0·865$ whatever be n.)

8. The joint-probability law of the y_i's for Example I of § 6 follows from an application of § 3. If we are considering a normal distribution and if $M(x) = \bar{x}$ then
$$y_i = \phi^{-1}\left(\frac{x_i - \bar{x}}{\sigma}\right).$$
Since the quantities $x_i - \bar{x}$ are independent of $\bar{x}$ it follows that the y's are also independent of $\bar{x}$. The most convenient formula to use would seem to be
$$p(y_1, \ldots, y_{n-1}, \bar{x} \mid \xi, \sigma) = \frac{p(x_1, \ldots, x_{n-1}, \bar{x} \mid \xi, \sigma)}{\prod_{i=1}^{n-1} f(x_i \mid \bar{x}, \sigma)}.$$
Since
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n-1} x_i + \frac{1}{n}x_n$$
and
$$\mathscr{E}\left(\frac{1}{n}x_n\right) = \frac{\xi}{n}, \quad \mathscr{E}\left(\left(\frac{1}{n}x_n\right)^2 - \left(\frac{\xi}{n}\right)^2\right)^{\frac{1}{2}} = \frac{\sigma}{n},$$
it is clear that
$$p(\bar{x} \mid x_1, \ldots, x_{n-1}, \xi, \sigma) = \frac{n}{\sqrt{(2\pi)}\,\sigma}\exp\left\{-\frac{1}{2\sigma^2}\left[(\bar{x}-\xi) - \sum_{i=1}^{n-1}(x_i - \bar{x})\right]^2\right\}.$$
Hence
$$p(x_1, \ldots, x_{n-1}, \bar{x} \mid \xi, \sigma) = \frac{n}{(\sqrt{(2\pi)}\,\sigma)^n}\exp\left(-\frac{1}{2\sigma^2}\left\{\sum_{i=1}^{n-1}(x_i-\xi)^2 + \left[(\bar{x}-\xi) - \sum_{i=1}^{n-1}(x_i-\bar{x})\right]^2\right\}\right),$$
and
$$\prod_{i=1}^{n-1} f(x_i \mid \bar{x}, \sigma) = \frac{1}{(\sqrt{(2\pi)}\,\sigma)^{n-1}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n-1}(x_i - \bar{x})^2\right].$$
The joint-probability law of $y_1, y_2, \ldots, y_{n-1}$ and $\bar{x}$ will be
$$p(y_1, \ldots, y_{n-1}, \bar{x} \mid \xi, \sigma) = \frac{n}{\sqrt{(2\pi)}\,\sigma}\exp\left(-\frac{1}{2\sigma^2}\left\{n(\bar{x}-\xi)^2 + \left[\sum_{i=1}^{n-1}(x_i - \bar{x})\right]^2\right\}\right)$$
$$= \frac{n}{\sqrt{(2\pi)}\,\sigma}\exp\left[-\frac{n(\bar{x}-\xi)^2}{2\sigma^2}\right]\exp\left[-\frac{1}{2}\left\{\sum_{i=1}^{n-1}\phi(y_i)\right\}^2\right],$$

$\phi(y_i)$ being defined as in (1). Integrate out for $\bar{x}$, and we have

$$p(y_1, \ldots, y_{n-1} \mid \xi, \sigma) = \sqrt{n} \exp\left[-\frac{1}{2}\left\{\sum_{i=1}^{n-1} \phi(y_i)\right\}^2\right]. \tag{6}$$

It will be noted that this joint probability law is independent of both ξ and σ.

9. The exponential law discussed in § 7 differs from the examples in which the normal law was used in that in this particular example a measure of location is used to estimate a scale parameter. We have

$$p(x) = \frac{1}{\theta} e^{-x/\theta},$$

whence, since

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n-1} x_i + \frac{1}{n} x_n,$$

it is seen that

$$p(\bar{x} \mid x_1, \ldots, x_{n-1}, \theta) = \frac{n}{\theta} \exp\left(-\frac{n}{\theta}\left[\bar{x} - \frac{1}{n}\sum_{i=1}^{n-1} x_i\right]\right),$$

and also

$$p(x_1, x_2, \ldots, x_{n-1} \mid \theta) = \frac{1}{\theta^{n-1}} \exp\left(-\frac{\sum_{i=1}^{n-1} x_i}{\theta}\right).$$

It follows that

$$p(x_1, \ldots, x_{n-1}, \bar{x} \mid \theta) = \frac{n}{\theta^n} \exp\left(-\frac{n\bar{x}}{\theta}\right),$$

and

$$\prod_{i=1}^{n-1} f(x_i \mid \bar{x}) = \frac{1}{\bar{x}^{n-1}} \exp\left(-\frac{\sum_{i=1}^{n-1} x_i}{\bar{x}}\right).$$

The joint-probability law of $y_1, \ldots, y_{n-1}$, and $\bar{x}$ follows in a straightforward way, namely

$$p(y_1, \ldots, y_{n-1}, \bar{x} \mid \theta) = \frac{n}{\theta}\left(\frac{\bar{x}}{\theta}\right)^{n-1} \exp\left(-\left[\frac{n\bar{x}}{\theta} - \frac{\sum_{i=1}^{n-1} x_i}{\bar{x}}\right]\right).$$

Remembering that

$$y_i = 1 - e^{-x_i/\bar{x}} \quad \text{or} \quad \frac{x_i}{\bar{x}} = -\log(1 - y_i)$$

the joint-probability law may be rewritten

$$p(y_1, \ldots, y_{n-1}, \bar{x} \mid \theta) = \frac{n}{\theta}\left(\frac{\bar{x}}{\theta}\right)^{n-1} e^{-n\bar{x}/\theta}\left[\frac{1}{\prod_{i=1}^{n-1}(1 - y_i)}\right],$$

where

$$\prod_{i=1}^{n-1}[-\log(1 - y_i)] < n \quad \text{or} \quad \prod_{i=1}^{n-1}(1 - y_i) < e^{-n}.$$

Integrating out with respect to $\bar{x}$,

$$p(y_1, \ldots, y_{n-1} \mid \theta) = \frac{(n-1)!}{(n-1)^{n-1}} \frac{1}{\prod_{i=1}^{n-1}(1 - y_i)}, \quad \prod_{i=1}^{n-1}(1 - y_i) < e^{-n}, \tag{7}$$

again a result which is independent of the parameter of the probability law.

10. The results of this investigation, which we have carried out partly in the general and partly in the particular, are obviously incomplete and should be succeeded by a fuller inquiry which would clear up the doubtful points which we have had to pass over, and possibly extend the general theory a little further. We feel that none of the questions which

have been raised in the course of this inquiry are insoluble by algebraic analysis but it is uncertain whether it is profitable to proceed with the fuller inquiry until some of the statistical implications of what has been done become more clear. For example, we have noted that given n independent random variables, x, if s sample moments are calculated from them and used as estimates of the parameters of the probability law, then it appears that there will be s independent relationships between the y's. Thus in this case the point $y_1, y_2, \ldots, y_n$ is constrained to move in an $n-s$ dimensioned space within an n dimensioned cube, and we have the exact analogue to the loss of degrees of freedom with χ^2 when the parameters have to be estimated from the data. What is not clear is how the y's are constrained when the sample estimates of the parameters are not the sample moments, and while this situation may not often be met with in practice, yet it should be explored.

When the parameters of location and scale are estimated from the data it is clear that the distribution of any individual y_i, and the joint-probability law of the y's also, will not be dependent on these unknown parameters of the probability law of the x's, but will depend on the functional form of that law. This result appears capable of extension for the case when higher sample moments are also used for estimating parameters. This being so, there would seem to be two ways in which the joint probability law of the y's may be utilized in statistical applications. First, it should be possible mathematically to form certain broad classes of functions for each of which the joint-probability laws of the y's would be approximately the same, or second one may seek for some transformation of variables so that instead of the correlated y_i we obtain $n-s$ new independent variables following some distributions which are independent of the original $p(x)$. Both these methods of attack may lead to results which will only be valid for large samples, but provided the results in either case have sufficient algebraic simplicity they should make possible certain generalizations in statistical analysis of which Neyman's 'smooth' test for goodness of fit is only one important example.

REFERENCES

Fisher, R. A. (1932). *Statistical Methods for Research Workers*, § 21·1.
Neyman, J. (1937). *Skand. AktuarTidskr.* **20**, 149.
Pearson, E. S. (1938). *Biometrika*, **30**, 134.
Pearson, K. (1933). *Biometrika*, **25**, 379.

A TABLE FOR THE CALCULATION OF WORKING PROBITS AND WEIGHTS IN PROBIT ANALYSIS

By D. J. FINNEY (*Lecturer in the Design and Analysis of Scientific Experiment, University of Oxford*) and W. L. STEVENS (*Admiralty*)

The estimation of the parameters of a distribution of individual tolerances, from data relating to numbers of subjects manifesting a characteristic quantal response at different levels of a stimulus, is a problem frequently encountered in the application of statistical science to dose-mortality studies, biological assay, detonation of explosives, and other problems. A typical situation is that of exposing batches of insects to various doses of an insecticide, recording the proportion killed at each level of dose and then requiring to estimate the mean tolerance (or median lethal dose) of individual insects and the variance of the tolerance distribution. Gaddum (1933) and Bliss (1935*a*, *b*; 1938) have been instrumental in developing a method, that of the *probit transformation*, which greatly simplifies the calculations necessary to the estimation. The exact statistical analysis appropriate to the transformation was first shown by Fisher (1935), and the theory and uses of the method have been discussed fully in many subsequent publications (Finney, 1947*a*, *b*).

Tables required in the practice of the method, in sufficient detail for most purposes, have been given by various writers (Fisher & Yates, 1943; Finney, 1947*a*). Occasionally, however, the statistician needs values of the various functions at finer intervals of the argument, and for his benefit the following Table has been prepared. A brief account of the tabulated functions will suffice for all who are familiar with the probit method; those who require fuller information on the theory and analysis should consult the list of References.

Given a proportion P, and its complement $Q = 1-P$, the probit of P is, to all intents and purposes, the deviate from the mean which divides the normal curve of unit variance in the ratio $P:Q$. In the formal definition, however, 5 is added to the deviate in order to avoid the necessity of computing with negative numbers. The advantage of this modification may be questioned, but it is now well established and will be adopted here. The probit, Y, of the proportion P is thus defined by

$$P = \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{Y-5} e^{-\frac{1}{2}u^2}\,du.$$

The standard method of analysis makes use of the *maximum and minimum working probits*,

$$Y_{\text{max.}} = Y + \frac{Q}{Z}$$

and

$$Y_{\text{min.}} = Y - \frac{P}{Z},$$

and also of the *range*,

$$1/Z,$$

where

$$Z = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}(Y-5)^2}.$$

If n subjects receive the same stimulus, and r of them show the characteristic response, the empirical value for the proportion responding is

$$p = r/n;$$

the complement of this is denoted by $q = 1-p$. The probits of a set of values of p should be approximately linearly related to x, the measure of the stimulus, and a line fitted by eye may be used to give a corresponding set of *expected probits*, Y. The *working probit* corresponding to each proportion is next calculated, from either

$$y = Y + Q/Z - q/Z,$$

or

$$y = Y - P/Z + p/Z,$$

using tabulated values of the maximum or the minimum working probit (whichever is the more convenient) and the range. An improved set of expected probits is then derived from the weighted linear regression equation of working probits on x, each y being assigned a weight, nw, where the *weighting coefficient*, w, is defined as

$$w = Z^2/PQ.$$

The process may be repeated with the new set of Y values. The iteration converges to give a linear regression equation which is an estimate of

$$Y = 5 + (x-\mu)/\sigma,$$

where μ is the mean and σ the standard deviation of the tolerance distribution. The method depends upon an assumption that the stimulus is measured on a scale for which individual tolerances are normally distributed: often the logarithm of 'dose' rather than dose itself is taken as x, in order to satisfy this condition more closely.

The Table which follows gives $Y_{\text{max.}}$ for $Y = 3{\cdot}58(0{\cdot}01)\ 9{\cdot}00$, $Y_{\text{min.}}$ for $Y = 1{\cdot}00(0{\cdot}01)\ 6{\cdot}42$, $1/Z$ and w for $Y = 1{\cdot}00(0{\cdot}01)\ 9{\cdot}00$, all to four places of decimals. Below $Y = 3{\cdot}58$, $Y_{\text{max.}}$ exceeds $10{\cdot}00$, and above $Y = 6{\cdot}52$, $Y_{\text{min.}}$ is negative; it is then almost always more convenient to calculate working probits from the other function, but the function not tabulated can easily be obtained from the relationship

$$Y_{\text{max.}} - Y_{\text{min.}} = 1/Z.$$

Between 3·58 and 6·42 both functions are tabulated. In order to save space the Table is arranged in parallel forward- and backward-reading columns; for the arguments Y and $(10-Y)$ values of $1/Z$ and w are the same, and simple relations exist between $Y_{\text{max.}}$ and $Y_{\text{min.}}$. All entries have been calculated to six or more places of decimals and rounded to four, except that w between $Y = 2{\cdot}7$ and $Y = 7{\cdot}3$ was obtained by collating two existing tables and checking discrepancies.

The values of P and Z, from which the present Table has been calculated, were taken from *Tables of the Probability Function*, Vol. II (1942), published by the Federal Works Project Administration for the City of New York. Values of Q/Z have been taken from, or checked against, W. F. Sheppard's table, published as *The Probability Integral* (1939), Vol. VII of the *British Association Mathematical Tables*.

Example

In a batch of 281 insects receiving the same dose of insecticide, 119 are killed. The provisional probit regression line gives an expected probit of 4·61 for this dose; find the working probit and the weight to be attached to the observation.

Expected probit Y	Maximum working probit $Y+Q/Z$	Minimum working probit $Y-P/Z$	Range $1/Z$	Weighting coefficient Z^2/PQ	$Y+Q/Z$ Maximum working probit	$Y-P/Z$ Minimum working probit	Y Expected probit
5·00	6·2533	3·7467	2·5066	0·6366	6·2533	3·7467	5·00
·01	·2534	·7466	·5068	·6366	·2534	·7466	4·99
·02	·2536	·7465	·5071	·6365	·2535	·7464	·98
·03	·2539	·7461	·5078	·6364	·2539	·7461	·97
·04	·2543	·7457	·5086	·6362	·2543	·7457	·96
5·05	6·2548	3·7450	2·5098	0·6360	6·2550	3·7452	4·95
·06	·2555	·7444	·5111	·6358	·2556	·7445	·94
·07	·2563	·7435	·5128	·6355	·2565	·7437	·93
·08	·2572	·7425	·5147	·6351	·2575	·7428	·92
·09	·2582	·7414	·5168	·6347	·2586	·7418	·91
5·10	6·2593	3·7401	2·5192	0·6343	6·2599	3·7407	4·90
·11	·2605	·7387	·5218	·6338	·2613	·7395	·89
·12	·2618	·7371	·5247	·6333	·2629	·7382	·88
·13	·2632	·7353	·5279	·6327	·2647	·7368	·87
·14	·2647	·7334	·5313	·6321	·2666	·7353	·86
5·15	6·2664	3·7314	2·5350	0·6314	6·2686	3·7336	4·85
·16	·2681	·7292	·5389	·6307	·2708	·7319	·84
·17	·2699	·7268	·5431	·6300	·2732	·7301	·83
·18	·2718	·7242	·5476	·6292	·2758	·7282	·82
·19	·2738	·7215	·5523	·6283	·2785	·7262	·81
5·20	6·2759	3·7186	2·5573	0·6274	6·2814	3·7241	4·80
·21	·2781	·7156	·5625	·6265	·2844	·7219	·79
·22	·2804	·7124	·5680	·6255	·2876	·7196	·78
·23	·2828	·7090	·5738	·6245	·2910	·7172	·77
·24	·2853	·7054	·5799	·6234	·2946	·7147	·76
5·25	6·2878	3·7016	2·5862	0·6223	6·2984	3·7122	4·75
·26	·2905	·6977	·5928	·6211	·3023	·7095	·74
·27	·2932	·6935	·5997	·6199	·3065	·7068	·73
·28	·2960	·6892	·6068	·6187	·3108	·7040	·72
·29	·2989	·6846	·6143	·6174	·3154	·7011	·71
5·30	6·3018	3·6798	2·6220	0·6161	6·3202	3·6982	4·70
·31	·3049	·6749	·6300	·6147	·3251	·6951	·69
·32	·3080	·6697	·6383	·6133	·3303	·6920	·68
·33	·3112	·6643	·6469	·6119	·3357	·6888	·67
·34	·3145	·6587	·6558	·6104	·3413	·6855	·66
5·35	6·3178	3·6528	2·6650	0·6088	6·3472	3·6822	4·65
·36	·3213	·6469	·6744	·6072	·3531	·6787	·64
·37	·3248	·6406	·6842	·6056	·3594	·6752	·63
·38	·3283	·6340	·6943	·6040	·3660	·6717	·62
·39	·3320	·6273	·7047	·6023	·3727	·6680	·61
5·40	6·3357	3·6203	2·7154	0·6005	6·3797	3·6643	4·60
·41	·3394	·6130	·7264	·5987	·3870	·6606	·59
·42	·3433	·6055	·7378	·5969	·3945	·6567	·58
·43	·3472	·5978	·7494	·5951	·4022	·6528	·57
·44	·3512	·5898	·7614	·5932	·4102	·6488	·56
5·45	6·3552	3·5815	2·7737	0·5912	6·4185	3·6448	4·55
·46	·3593	·5729	·7864	·5893	·4271	·6407	·54
·47	·3635	·5641	·7994	·5872	·4359	·6365	·53
·48	·3677	·5550	·8127	·5852	·4450	·6323	·52
·49	·3720	·5456	·8264	·5831	·4544	·6280	·51
5·50	6·3764	3·5360	2·8404	0·5810	6·4640	3·6236	4·50

Expected probit Y	Maximum working probit $Y+Q/Z$	Minimum working probit $Y-P/Z$	Range $1/Z$	Weighting coefficient Z^2/PQ			
5·50	6·3764	3·5360	2·8404	0·5810	6·4640	3·6236	4·50
·51	·3808	·5260	·8548	·5788	·4740	·6192	·49
·52	·3852	·5157	·8695	·5766	·4843	·6148	·48
·53	·3898	·5052	·8846	·5744	·4948	·6102	·47
·54	·3944	·4943	·9001	·5722	·5057	·6056	·46
5·55	6·3990	3·4831	2·9159	0·5699	6·5169	3·6010	4·45
·56	·4037	·4715	·9322	·5675	·5285	·5963	·44
·57	·4085	·4597	·9488	·5652	·5403	·5915	·43
·58	·4133	·4475	·9658	·5628	·5525	·5867	·42
·59	·4181	·4349	·9832	·5603	·5651	·5819	·41
5·60	6·4230	3·4220	3·0010	0·5579	6·5780	3·5770	4·40
·61	·4280	·4088	·0192	·5554	·5912	·5720	·39
·62	·4330	·3952	·0378	·5529	·6048	·5670	·38
·63	·4381	·3812	·0569	·5503	·6188	·5619	·37
·64	·4432	·3669	·0763	·5477	·6331	·5568	·36
5·65	6·4484	3·3522	3·0962	0·5451	6·6478	3·5516	4·35
·66	·4536	·3370	·1166	·5425	·6630	·5464	·34
·67	·4588	·3214	·1374	·5398	·6786	·5412	·33
·68	·4641	·3055	·1586	·5371	·6945	·5359	·32
·69	·4695	·2892	·1803	·5343	·7108	·5305	·31
5·70	6·4749	3·2724	3·2025	0·5316	6·7276	3·5251	4·30
·71	·4803	·2551	·2252	·5288	·7449	·5197	·29
·72	·4858	·2375	·2483	·5260	·7625	·5142	·28
·73	·4914	·2194	·2720	·5232	·7806	·5086	·27
·74	·4969	·2008	·2961	·5203	·7992	·5031	·26
5·75	6·5026	3·1819	3·3207	0·5174	6·8181	3·4974	4·25
·76	·5082	·1623	·3459	·5145	·8377	·4918	·24
·77	·5139	·1423	·3716	·5116	·8577	·4861	·23
·78	·5197	·1219	·3978	·5086	·8781	·4803	·22
·79	·5255	·1009	·4246	·5056	·8991	·4745	·21
5·80	6·5313	3·0794	3·4519	0·5026	6·9206	3·4687	4·20
·81	·5372	·0574	·4798	·4996	·9426	·4628	·19
·82	·5431	·0348	·5083	·4965	·9652	·4569	·18
·83	·5490	·0116	·5374	·4935	·9884	·4510	·17
·84	·5550	2·9880	·5670	·4904	7·0120	·4450	·16
5·85	6·5611	2·9638	3·5973	0·4873	7·0362	3·4389	4·15
·86	·5671	·9389	·6282	·4841	·0611	·4329	·14
·87	·5732	·9135	·6597	·4810	·0865	·4268	·13
·88	·5794	·8875	·6919	·4778	·1125	·4206	·12
·89	·5855	·8608	·7247	·4746	·1392	·4145	·11
5·90	6·5917	2·8335	3·7582	0·4714	7·1665	3·4083	4·10
·91	·5980	·8056	·7924	·4682	·1944	·4020	·09
·92	·6043	·7771	·8272	·4650	·2229	·3957	·08
·93	·6106	·7478	·8628	·4617	·2522	·3894	·07
·94	·6169	·7178	·8991	·4585	·2822	·3831	·06
5·95	6·6233	2·6872	3·9361	0·4552	7·3128	3·3767	4·05
·96	·6297	·6558	·9739	·4519	·3442	·3703	·04
·97	·6362	·6238	4·0124	·4486	·3762	·3638	·03
·98	·6426	·5909	·0517	·4453	·4091	·3574	·02
·99	·6491	·5573	·0918	·4420	·4427	·3509	·01
6·00	6·6557	2·5230	4·1327	0·4386	7·4770	3·3443	4·00
			$1/Z$ Range	Z^2/PQ Weighting coefficient	$Y+Q/Z$ Maximum working probit	$Y-P/Z$ Minimum working probit	Y Expected probit

Expected probit Y	Maximum working probit $Y+Q/Z$	Minimum working probit $Y-P/Z$	Range $1/Z$	Weighting coefficient Z^2/PQ			
6·00	6·6557	2·5230	4·1327	0·4386	7·4770	3·3443	4·00
·01	·6623	·4878	·1745	·4353	·5122	·3377	3·99
·02	·6689	·4518	·2171	·4319	·5482	·3311	·98
·03	·6755	·4150	·2605	·4285	·5850	·3245	·97
·04	·6822	·3774	·3048	·4252	·6226	·3178	·96
6·05	6·6888	2·3387	4·3501	0·4218	7·6613	3·3112	3·95
·06	·6956	·2994	·3962	·4184	·7006	·3044	·94
·07	·7023	·2590	·4433	·4150	·7410	·2977	·93
·08	·7091	·2178	·4913	·4116	·7822	·2909	·92
·09	·7159	·1756	·5403	·4082	·8244	·2841	·91
6·10	6·7227	2·1324	4·5903	0·4047	7·8676	3·2773	3·90
·11	·7296	·0883	·6413	·4013	·9117	·2704	·89
·12	·7365	·0432	·6933	·3979	·9568	·2635	·88
·13	·7434	1·9970	·7464	·3944	8·0030	·2566	·87
·14	·7504	·9498	·8006	·3910	·0502	·2496	·86
6·15	6·7573	1·9014	4·8559	0·3876	8·0986	3·2427	3·85
·16	·7643	·8520	·9123	·3841	·1480	·2357	·84
·17	·7714	·8016	·9698	·3807	·1984	·2286	·83
·18	·7784	·7498	5·0286	·3772	·2502	·2216	·82
·19	·7855	·6970	·0885	·3738	·3030	·2145	·81
6·20	6·7926	1·6429	5·1497	0·3703	8·3571	3·2074	3·80
·21	·7997	·5876	·2121	·3669	·4124	·2003	·79
·22	·8068	·5310	·2758	·3634	·4690	·1932	·78
·23	·8140	·4731	·3409	·3600	·5269	·1860	·77
·24	·8212	·4140	·4072	·3565	·5860	·1788	·76
6·25	6·8284	1·3534	5·4750	0·3531	8·6466	3·1716	3·75
·26	·8357	·2916	·5441	·3496	·7084	·1643	·74
·27	·8429	·2282	·6147	·3462	·7718	·1571	·73
·28	·8502	·1635	·6867	·3428	·8365	·1498	·72
·29	·8575	·0972	·7603	·3393	·9028	·1425	·71
6·30	6·8649	1·0295	5·8354	0·3359	8·9705	3·1351	3·70
·31	·8722	0·9602	·9120	·3325	9·0398	·1278	·69
·32	·8796	·8893	·9903	·3291	·1107	·1204	·68
·33	·8870	·8168	6·0702	·3256	·1832	·1130	·67
·34	·8944	·7426	·1518	·3222	·2574	·1056	·66
6·35	6·9019	0·6668	6·2351	0·3188	9·3332	3·0981	3·65
·36	·9093	·5892	·3201	·3155	·4108	·0907	·64
·37	·9168	·5098	·4070	·3121	·4902	·0832	·63
·38	·9243	·4286	·4957	·3087	·5714	·0757	·62
·39	·9318	·3455	·5863	·3053	·6545	·0682	·61
6·40	6·9394	0·2606	6·6788	0·3020	9·7394	3·0606	3·60
·41	·9469	·1736	·7733	·2986	·8264	·0531	·59
·42	·9545	·0847	·8698	·2953	·9153	·0455	·58
·43	·9621		·9684	·2920		·0379	·57
·44	·9697		7·0691	·2887		·0303	·56
6·45	6·9774		7·1720	0·2854		3·0226	3·55
·46	·9850		·2771	·2821		·0150	·54
·47	6·9927		·3845	·2788		·0073	·53
·48	7·0004		·4943	·2756		2·9996	·52
·49	·0081		·6064	·2723		·9919	·51
6·50	7·0158		7·7210	0·2691		2·9842	3·50
			$1/Z$ Range	Z^2/PQ Weighting coefficient	$Y+Q/Z$ Maximum working probit	$Y-P/Z$ Minimum working probit	Y Expected probit

Expected probit Y	Maximum working probit $Y+Q/Z$	Range $1/Z$	Weighting coefficient Z^2/PQ		
6·50	7·0158	7·7210	0·2691	2·9842	3·50
·51	·0236	·8380	·2658	·9764	·49
·52	·0313	·9577	·2626	·9687	·48
·53	·0391	8·0800	·2594	·9609	·47
·54	·0469	·2050	·2563	·9531	·46
6·55	7·0547	8·3327	0·2531	2·9453	3·45
·56	·0625	·4633	·2500	·9375	·44
·57	·0704	·5968	·2468	·9296	·43
·58	·0783	·7333	·2437	·9217	·42
·59	·0861	·8728	·2406	·9139	·41
6·60	7·0940	9·0154	0·2375	2·9060	3·40
·61	·1020	·1613	·2345	·8980	·39
·62	·1099	·3105	·2314	·8901	·38
·63	·1178	·4630	·2284	·8822	·37
·64	·1258	·6190	·2254	·8742	·36
6·65	7·1338	9·7785	0·2224	2·8662	3·35
·66	·1417	·9417	·2194	·8583	·34
·67	·1498	10·1086	·2165	·8502	·33
·68	·1578	10·2794	·2135	·8422	·32
·69	·1658	10·4540	·2106	·8342	·31
6·70	7·1739	10·6327	0·2077	2·8261	3·30
·71	·1819	10·8156	·2049	·8181	·29
·72	·1900	11·0027	·2020	·8100	·28
·73	·1981	11·1941	·1992	·8019	·27
·74	·2062	11·3900	·1964	·7938	·26
6·75	7·2143	11·5905	0·1936	2·7857	3·25
·76	·2224	11·7957	·1908	·7776	·24
·77	·2306	12·0058	·1881	·7694	·23
·78	·2387	12·2208	·1853	·7613	·22
·79	·2469	12·4409	·1826	·7531	·21
6·80	7·2551	12·6662	0·1799	2·7449	3·20
·81	·2633	12·8969	·1773	·7367	·19
·82	·2715	13·1331	·1746	·7285	·18
·83	·2797	13·3750	·1720	·7203	·17
·84	·2880	13·6227	·1694	·7120	·16
6·85	7·2962	13·8764	0·1669	2·7038	3·15
·86	·3045	14·1362	·1643	·6955	·14
·87	·3128	14·4023	·1618	·6872	·13
·88	·3210	14·6749	·1593	·6790	·12
·89	·3293	14·9541	·1568	·6707	·11
6·90	7·3376	15·2402	0·1544	2·6624	3·10
·91	·3460	15·5333	·1519	·6540	·09
·92	·3543	15·8337	·1495	·6457	·08
·93	·3626	16·1414	·1471	·6374	·07
·94	·3710	16·4568	·1448	·6290	·06
6·95	7·3794	16·7800	0·1424	2·6206	3·05
·96	·3877	17·1113	·1401	·6123	·04
·97	·3961	17·4509	·1378	·6039	·03
·98	·4045	17·7989	·1356	·5955	·02
·99	·4129	18·1558	·1333	·5871	·01
7·00	7·4214	18·5216	0·1311	2·5786	3·00
		$1/Z$ Range	Z^2/PQ Weighting coefficient	$Y-P/Z$ Minimum working probit	Y Expected probit

Expected probit Y	Maximum working probit $Y+Q/Z$	Range $1/Z$	Weighting coefficient Z^2/PQ		
7·00	7·4214	18·5216	0·1311	2·5786	3·00
·01	·4298	18·8967	·1289	·5702	2·99
·02	·4382	19·2814	·1268	·5618	·98
·03	·4467	19·6758	·1246	·5533	·97
·04	·4552	20·0803	·1225	·5448	·96
7·05	7·4636	20·4952	0·1204	2·5364	2·95
·06	·4721	20·9207	·1183	·5279	·94
·07	·4806	21·3572	·1163	·5194	·93
·08	·4891	21·8050	·1142	·5109	·92
·09	·4976	22·2644	·1122	·5024	·91
7·10	7·5062	22·7357	0·1103	2·4938	2·90
·11	·5147	23·2194	·1083	·4853	·89
·12	·5232	23·7157	·1064	·4768	·88
·13	·5318	24·2251	·1045	·4682	·87
·14	·5404	24·7478	·1026	·4596	·86
7·15	7·5489	25·2844	0·1007	2·4511	2·85
·16	·5575	25·8352	·0989	·4425	·84
·17	·5661	26·4006	·0971	·4339	·83
·18	·5747	26·9812	·0953	·4253	·82
·19	·5833	27·5772	·0935	·4167	·81
7·20	7·5919	28·1892	0·0918	2·4081	2·80
·21	·6006	28·8177	·0901	·3994	·79
·22	·6092	29·4631	·0884	·3908	·78
·23	·6178	30·1260	·0867	·3822	·77
·24	·6265	30·8069	·0851	·3735	·76
7·25	7·6351	31·5063	0·0834	2·3649	2·75
·26	·6438	32·2249	·0818	·3562	·74
·27	·6525	32·9631	·0802	·3475	·73
·28	·6612	33·7216	·0787	·3388	·72
·29	·6699	34·5010	·0771	·3301	·71
7·30	7·6786	35·3020	0·0756	2·3214	2·70
·31	·6873	36·1251	·0741	·3127	·69
·32	·6960	36·9712	·0727	·3040	·68
·33	·7047	37·8408	·0712	·2953	·67
·34	·7135	38·7348	·0698	·2865	·66
7·35	7·7222	39·6539	0·0684	2·2778	2·65
·36	·7310	40·5988	·0671	·2690	·64
·37	·7397	41·5704	·0656	·2603	·63
·38	·7485	42·5695	·0643	·2515	·62
·39	·7573	43·5970	·0630	·2427	·61
7·40	7·7661	44·6538	0·0617	2·2339	2·60
·41	·7748	45·7407	·0604	·2252	·59
·42	·7836	46·8588	·0591	·2164	·58
·43	·7924	48·0090	·0579	·2076	·57
·44	·8013	49·1924	·0567	·1987	·56
7·45	7·8101	50·4099	0·0555	2·1899	2·55
·46	·8189	51·6628	·0543	·1811	·54
·47	·8277	52·9521	·0532	·1723	·53
·48	·8366	54·2791	·0520	·1634	·52
·49	·8454	55·6448	·0509	·1546	·51
7·50	7·8543	57·0506	0·0498	2·1457	2·50
		$1/Z$ Range	Z^2/PQ Weighting coefficient	$Y-P/Z$ Minimum working probit	Y Expected probit

Expected probit Y	Maximum working probit $Y+Q/Z$	Range $1/Z$	Weighting coefficient Z^2/PQ		
7·50	7·8543	57·0506	0·0498	2·1457	2·50
·51	·8631	58·4978	·0487	·1369	·49
·52	·8720	59·9876	·0476	·1280	·48
·53	·8809	61·5216	·0466	·1191	·47
·54	·8897	63·1011	·0456	·1103	·46
7·55	7·8986	64·7277	0·0446	2·1014	2·45
·56	·9075	66·4028	·0436	·0925	·44
·57	·9164	68·1280	·0426	·0836	·43
·58	·9253	69·9051	·0416	·0747	·42
·59	·9342	71·7357	·0407	·0658	·41
7·60	7·9432	73·6216	0·0398	2·0568	2·40
·61	·9521	75·5646	·0389	·0479	·39
·62	·9610	77·5667	·0380	·0390	·38
·63	·9700	79·6298	·0371	·0300	·37
·64	·9789	81·7559	·0362	·0211	·36
7·65	7·9879	83·9472	0·0354	2·0121	2·35
·66	·9968	86·2059	·0346	·0032	·34
·67	8·0058	88·5342	·0338	1·9942	·33
·68	·0147	90·9344	·0330	·9853	·32
·69	·0237	93·4091	·0322	·9763	·31
7·70	8·0327	95·9607	0·0314	1·9673	2·30
·71	·0417	98·5918	·0307	·9583	·29
·72	·0507	101·3053	·0300	·9493	·28
·73	·0597	104·1038	·0292	·9403	·27
·74	·0687	106·9903	·0285	·9313	·26
7·75	8·0777	109·9679	0·0278	1·9223	2·25
·76	·0867	113·0396	·0272	·9133	·24
·77	·0957	116·2088	·0265	·9043	·23
·78	·1047	119·4788	·0258	·8953	·22
·79	·1138	122·8530	·0252	·8862	·21
7·80	8·1228	126·3352	0·0246	1·8772	2·20
·81	·1318	129·9290	·0240	·8682	·19
·82	·1409	133·6385	·0234	·8591	·18
·83	·1499	137·4676	·0228	·8501	·17
·84	·1590	141·4206	·0222	·8410	·16
7·85	8·1681	145·5018	0·0217	1·8319	2·15
·86	·1771	149·7158	·0211	·8229	·14
·87	·1862	154·0671	·0206	·8138	·13
·88	·1953	158·5609	·0200	·8047	·12
·89	·2044	163·2020	·0195	·7956	·11
7·90	8·2134	167·9957	0·0190	1·7866	2·10
·91	·2225	172·9476	·0185	·7775	·09
·92	·2316	178·0632	·0181	·7684	·08
·93	·2407	183·3485	·0176	·7593	·07
·94	·2498	188·8095	·0171	·7502	·06
7·95	8·2590	194·4526	0·0167	1·7410	2·05
·96	·2681	200·2844	·0162	·7319	·04
·97	·2772	206·3118	·0158	·7228	·03
·98	·2863	212·5418	·0154	·7137	·02
·99	·2955	218·9818	·0150	·7045	·01
8·00	8·3046	225·6395	0·0146	1·6954	2·00
		$1/Z$ Range	Z^2/PQ Weighting coefficient	$Y-P/Z$ Minimum working probit	Y Expected probit

Expected probit Y	Maximum working probit $Y+Q/Z$	Range $1/Z$	Weighting coefficient Z^2/PQ		
8·00	8·3046	225·6395	0·0146	1·6954	2·00
·01	·3137	232·5229	·0142	·6863	1·99
·02	·3229	239·6402	·0138	·6771	·98
·03	·3320	247·0000	·0134	·6680	·97
·04	·3412	254·6114	·0131	·6588	·96
8·05	8·3503	262·4836	0·0127	1·6497	1·95
·06	·3595	270·6262	·0124	·6405	·94
·07	·3687	279·0493	·0120	·6313	·93
·08	·3778	287·7634	·0117	·6222	·92
·09	·3870	296·7792	·0114	·6130	·91
8·10	8·3962	306·1082	0·0110	1·6038	1·90
·11	·4054	315·7619	·0107	·5946	·89
·12	·4146	325·7527	·0104	·5854	·88
·13	·4238	336·0932	·0101	·5762	·87
·14	·4330	346·7966	·0099	·5670	·86
8·15	8·4422	357·8732	0·0096	1·5578	1·85
·16	·4514	369·3477	·0093	·5486	·84
·17	·4606	381·2245	·0090	·5394	·83
·18	·4698	393·5226	·0088	·5302	·82
·19	·4790	406·2580	·0085	·5210	·81
8·20	8·4882	419·4476	0·0083	1·5118	1·80
·21	·4974	433·1086	·0080	·5026	·79
·22	·5067	447·2593	·0078	·4933	·78
·23	·5159	461·9185	·0076	·4841	·77
·24	·5251	477·1059	·0074	·4749	·76
8·25	8·5344	492·8419	0·0071	1·4656	1·75
·26	·5436	509·1479	·0069	·4564	·74
·27	·5529	526·0459	·0067	·4471	·73
·28	·5621	543·5592	·0065	·4379	·72
·29	·5714	561·7116	·0063	·4286	·71
8·30	8·5806	580·5283	0·0061	1·4194	1·70
·31	·5899	600·0353	·0060	·4101	·69
·32	·5992	620·2599	·0058	·4008	·68
·33	·6084	641·2302	·0056	·3916	·67
·34	·6177	662·9758	·0054	·3823	·66
8·35	8·6270	685·5274	0·0053	1·3730	1·65
·36	·6363	708·9171	·0051	·3637	·64
·37	·6456	733·1780	·0050	·3544	·63
·38	·6548	758·3451	·0048	·3452	·62
·39	·6641	784·4545	·0047	·3359	·61
8·40	8·6734	811·5439	0·0045	1·3266	1·60
·41	·6827	839·6528	·0044	·3173	·59
·42	·6920	868·8222	·0042	·3080	·58
·43	·7013	899·0948	·0041	·2987	·57
·44	·7106	930·5153	·0040	·2894	·56
8·45	8·7200	963·1301	0·0038	1·2800	1·55
·46	·7293	996·9878	·0037	·2707	·54
·47	·7386	1032·1389	·0036	·2614	·53
·48	·7479	1068·6362	·0035	·2521	·52
·49	·7572	1106·5347	·0034	·2428	·51
8·50	8·7666	1145·8919	0·0033	1·2334	1·50
		$1/Z$ Range	Z^2/PQ Weighting coefficient	$Y-P/Z$ Minimum working probit	Y Expected probit

Expected probit Y	Maximum working probit $Y+Q/Z$	Range $1/Z$	Weighting coefficient Z^2/PQ	Minimum working probit $Y-P/Z$	Expected probit Y
8·50	8·7666	1145·8919	0·0033	1·2334	1·50
·51	·7759	1186·7675	·0032	·2241	·49
·52	·7852	1229·2242	·0031	·2148	·48
·53	·7946	1273·3271	·0030	·2054	·47
·54	·8039	1319·1443	·0029	·1961	·46
8·55	8·8133	1366·7467	0·0028	1·1867	1·45
·56	·8226	1416·2085	·0027	·1774	·44
·57	·8320	1467·6071	·0026	·1680	·43
·58	·8413	1521·0232	·0025	·1587	·42
·59	·8507	1576·5411	·0024	·1493	·41
8·60	8·8600	1634·2488	0·0024	1·1400	1·40
·61	·8694	1694·2383	·0023	·1306	·39
·62	·8788	1756·6055	·0022	·1212	·38
·63	·8881	1821·4507	·0021	·1119	·37
·64	·8975	1888·8785	·0021	·1025	·36
8·65	8·9069	1958·9983	0·0020	1·0931	1·35
·66	·9162	2031·9243	·0019	·0838	·34
·67	·9256	2107·7758	·0019	·0744	·33
·68	·9350	2186·6775	·0018	·0650	·32
·69	·9444	2268·7596	·0017	·0556	·31
8·70	8·9538	2354·1583	0·0017	1·0462	1·30
·71	·9632	2443·0158	·0016	·0368	·29
·72	·9726	2535·4807	·0016	·0274	·28
·73	·9820	2631·7085	·0015	·0180	·27
·74	·9914	2731·8615	·0015	·0086	·26
8·75	9·0008	2836·1096	0·0014	0·9992	1·25
·76	·0102	2944·6302	·0014	·9898	·24
·77	·0196	3057·6091	·0013	·9804	·23
·78	·0290	3175·2401	·0013	·9710	·22
·79	·0384	3297·7264	·0012	·9616	·21
8·80	9·0478	3425·2801	0·0012	0·9522	1·20
·81	·0572	3558·1233	·0011	·9428	·19
·82	·0667	3696·4883	·0011	·9333	·18
·83	·0761	3840·6179	·0011	·9239	·17
·84	·0855	3990·7662	·0010	·9145	·16
8·85	9·0949	4147·1994	0·0010	0·9051	1·15
·86	·1044	4310·1955	·0010	·8956	·14
·87	·1138	4480·0457	·0009	·8862	·13
·88	·1232	4657·0549	·0009	·8768	·12
·89	·1327	4841·5419	·0009	·8673	·11
8·90	9·1421	5033·8407	0·0008	0·8579	1·10
·91	·1516	5234·3007	·0008	·8484	·09
·92	·1610	5443·2878	·0008	·8390	·08
·93	·1704	5661·1851	·0007	·8296	·07
·94	·1799	5888·3938	·0007	·8201	·06
8·95	9·1894	6125·3338	0·0007	0·8106	1·05
·96	·1988	6372·4452	·0007	·8012	·04
·97	·2083	6630·1886	·0006	·7917	·03
·98	·2177	6899·0468	·0006	·7823	·02
·99	·2272	7179·5252	·0006	·7728	·01
9·00	9·2367	7472·1536	0·0006	0·7633	1·00

The proportion killed is $p = 119/281 = 0{\cdot}4235.$

For $Y = 4{\cdot}61$, the Table shows the minimum working probit and range as

$$Y_{\text{min.}} = 3{\cdot}6680,$$

and $$1/Z = 2{\cdot}7047.$$

Hence the working probit is

$$Y = 3{\cdot}6680 + 2{\cdot}7047p$$
$$= 4{\cdot}8134.$$

Alternatively, if survivors instead of deaths have been recorded, calculation may proceed from

$$Y_{\text{max.}} = 6{\cdot}3727,$$

giving

$$y = 6{\cdot}3727 - 2{\cdot}7047 \times 0{\cdot}5765$$
$$= 4{\cdot}8134.$$

The Table shows $w = 0{\cdot}6023,$

so that the weight for the observation is

$$nw = 281 \times 0{\cdot}6023$$
$$= 169{\cdot}2.$$

REFERENCES

BLISS, C. I. (1935*a*). The calculation of the dosage-mortality curve. *Ann. Appl. Biol.* **22**, 134–67.

BLISS, C. I. (1935*b*). The comparison of dosage mortality data. *Ann. Appl. Biol.* **22**, 307–33.

BLISS, C. I. (1938). The determination of dosage-mortality curves from small numbers. *Quart. J. Pharm.* **11**, 192–216.

FINNEY, D. J. (1947*a*). *Probit Analysis: A Statistical Treatment of the Sigmoid Response Curve.* London: Cambridge University Press.

FINNEY, D. J. (1947*b*). The principles of biological assay. *J. Roy. Statist. Soc. Suppl.* **9**, 46–91.

FISHER, R. A. (1935). Appendix to BLISS, C. I.: The case of zero survivors. *Ann. Appl. Biol.* **22**, 164–5.

FISHER, R. A. & YATES, F. (1943). *Statistical Tables for Biological, Agricultural and Medical Research* (2nd ed.). Edinburgh: Oliver and Boyd.

GADDUM, J. H. (1933). Reports on biological standards. III. Methods of biological assay depending on a quantal response. *Spec. Rep. Ser. Med. Res. Coun., Lond.*, no. 183.

MISCELLANEA

A note on the χ^2 smooth test

By H. L. Seal

The now-classical test of goodness of fit of a theoretical frequency distribution to a set of observations consists of the calculation of $\sum_{j=1}^{n} (\delta m_j)^2/m_j$ (where m_j is the expectation of the jth group of observations and δm_j the deviation of the actual number of observations in this group from its expectation) and reference to tables of χ^2 with $n-1$ degrees of freedom. This is under the assumption that the only restraint on the theoretical frequency curve is that $\sum_{j=1}^{n} \delta m_j = 0$. In her recent contribution to this *Journal*, F. N. David (1947) shows by a geometrical method that, under these circumstances, a test of the significance of sequences of like signs in $\delta m_j (j = 1, 2, \ldots, n)$ is, for practical purposes, stochastically independent of the χ^2-test. The restriction of the arguments to cover only the case of a single linear restraint is, however, unnecessary.

It is possible, in this connexion, to state a useful general theorem: *If $x_j (j = 1, 2, \ldots, n)$ are n random variables normally distributed about zero mean with unit variance, these variables being connected by means of k linear relations, the probability distribution of $q^2 = \sum_{j=1}^{n} x_j^2$ remains unaltered if we select only those samples in which the signs of x_j follow a specified pattern.*

The proof is immediate since

$$p\{x_1, x_2, \ldots, x_n\} = \frac{1}{2^{\frac{1}{2}(n-k)-1}\,\Gamma[\frac{1}{2}(n-k)]} e^{-\frac{1}{2}q^2} q^{n-k-1} F(u_1, u_2, \ldots, u_{n-1}),$$

where $x_j = qu_j \quad j = 1, 2, \ldots, n, \qquad \sum_{j=1}^{n} u_j^2 = 1$

and the $u_j (j = 1, 2, \ldots, n-1)$ are connected by k further relations (cp. Hald & Rasch, 1943). Thus, changes in the signs of the x's are reflected by changes only in the signs of the u's, the joint distribution of which is independent of that of q.

Naturally this theorem does not apply directly to the case considered by David since the χ^2-test is there only an approximation to a set of terms of a multinomial expansion. It is, moreover, easily seen that if the squares of n independent random variables, with the same arbitrary skew distribution law, are added, the probability distribution of the resulting sum will necessarily depend on the signs of these variables. However, the practical use of the χ^2-test as an approximation implies 'near independence' of the test for sequences of signs even when k parameters (instead of one) have been fitted and have reduced the degrees of freedom to $n-k$.

A further point may be made. David's method of combining the χ^2-test of goodness of fit with the sequence test envisages a frequency distribution $m_j (j = 1, 2, \ldots, n)$ where n is relatively small: in fact her tables for the application of the combined test are calculated for $n = 5(1)\,14$. There is, however, a closely analogous case where n may assume a value between about 30 and 75, namely in testing the efficacy of the graduation of a mortality table. The writer (1943) suggested that in this case a good test would consist of the application of a χ^2-test of goodness of fit with $n-k$ degrees of freedom and a sequence test in the form of a 2×2 contingency table and associated χ^2 value, as indicated by Stevens (1939) himself when he suggested this latter test. In answer to an inquiry, the writer said he was 'not sure of the complete independence of the tests' but in view of the preceding it may be said that the second χ^2 (one degree of freedom) is very nearly independent of the χ^2 value in the main test, so that the two can be conveniently combined by addition into one value with $n-k+1$ degrees of freedom and a useful probability judgement of 'smooth fit' obtained.

REFERENCES

David, F. N. (1947). A χ^2 'smooth' test for goodness of fit. *Biometrika*, **34**, 299.

Hald, A. & Rasch, G. (1943). Nogle Anvendelser af Transformationsmetoden i den normale Fordelings Teori. *Festskrift til Prof. J. F. Steffensen.* Copenhagen.

Seal, H. L. (1943). Tests of a mortality table graduation. *J. Inst. Actu.* **71**, 5.

Stevens, W. L. (1939). Distribution of groups in a sequence of alternatives. *Ann. Eugen., Lond.*, **9**, 10.

Rank correlation and product-moment correlation

By P. A. P. MORAN, *Institute of Statistics, Oxford University*

The sampling distribution of Spearman's coefficient of rank correlation, ρ_s, has been thoroughly studied in the case where every permutation of the ranks of one variate relative to another is equiprobable and more recently Höffding (1948) has shown that it tends to normality for large samples whatever the parent population. This note deals with the distribution, when the parent is normal, for any size of sample.

Let $(x_1, y_1) \ldots (x_n, y_n)$ be a sample of n pairs of values from a bivariate normal population with correlation coefficient ρ. To obtain ρ_s we replace $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ by their respective ranks and calculate the product moment correlation coefficient of the ranks as if they were variate values. We have then to discuss the distribution of ρ_s when ρ is known.

Consider first the expected value of ρ_s. Let $p(x_i), p(y_i)$ be the ranks of x_i and y_i. Write

$$H(t) = 0 \quad \text{for} \quad t \leqslant 0$$
$$= 1 \qquad t > 0.$$

Then
$$p(x_i) - 1 = \sum_{j=1}^{n} H(x_i - x_j). \tag{1}$$

ρ_s will be the correlation coefficient of the numbers $\{p(x_i)-1\}$ and $\{p(y_i)-1\}$. If we had defined $H(t)$ in such a way that $H(0) = 1$ we would have obtained $p(x_i)$ on the left-hand side of equation (1) but this greatly complicates later calculations. Now write

$$S = \sum_{i=1}^{n} (p(x_i)-1)(p(y_i)-1).$$

It is easy to verify that
$$\sum_{i=1}^{n} \{p(x_i)-1\} = \tfrac{1}{2}n(n-1),$$

and
$$\sum_{i=1}^{n} \{p(x_i)-\bar{p}\}^2 = \tfrac{1}{12}n(n^2-1),$$

where $\bar{p}$ is the mean of the ranks.

It follows that
$$\rho_s = \frac{s - \frac{1}{4}n(n-1)^2}{\frac{1}{12}n(n^2-1)},$$

and to find $E(\rho_s)$ it is enough to find $E(s)$. Now

$$s = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} H(x_i - x_j) H(y_i - y_k).$$

The terms in this expansion for which $i = j$ or $i = k$ are zero. There remain only two cases to consider.

Case I. i, j, k all distinct. Then $x_i - x_j, y_i - y_k$ are distributed in a bivariate normal distribution with correlation coefficient equal to $\frac{1}{2}\rho$. $E\{H(x_i - x_j) H(y_i - y_k)\}$ will therefore be the chance that both these quantities are positive, that is, the integral of the probability density over the positive quadrant. This is known (Sheppard, 1898) to be equal to $\frac{1}{2}\{1 - \pi^{-1}\cos^{-1}\frac{1}{2}\rho\}$. There are clearly $n(n-1)(n-2)$ such terms.

Case II. $i \neq j = k$. Then $x_i - x_j$ and $y_i - y_k$ have a correlation coefficient ρ and the expectation of each term is $\frac{1}{2}\{1 - \pi^{-1}\cos^{-1}\rho\}$. The number of such terms is $n(n-1)$.

It follows that

$$E(s) = \tfrac{1}{2}n(n-1)(n-2)(1-\pi^{-1}\cos^{-1}\tfrac{1}{2}\rho) + \tfrac{1}{2}n(n-1)(1-\pi^{-1}\cos^{-1}\rho),$$

and so
$$E(\rho_s) = \frac{12}{n(n^2-1)}\{\tfrac{1}{2}n(n-1)(n-2)(1-\pi^{-1}\cos^{-1}\tfrac{1}{2}\rho)$$
$$+ \tfrac{1}{2}n(n-1)(1-\pi^{-1}\cos^{-1}\rho) - \tfrac{1}{4}n(n-1)^2\}$$
$$= \frac{6}{\pi}\left\{\frac{n-2}{n+1}\sin^{-1}\tfrac{1}{2}\rho + \frac{1}{n+1}\sin^{-1}\rho\right\}. \tag{2}$$

For $\rho = 0, 1$ this equals zero or unity as we would expect. Equation (2) may be compared with K. Pearson's approximate formula (1907) for turning rank correlation coefficients into product moment correlation coefficients, which he derived by using the correlation of grades. This is

$$\rho_s = \frac{6}{\pi}\sin^{-1}\tfrac{1}{2}\rho.$$

Equation (2) tends to this as n increases, showing that ρ_s is not only a biased estimator of ρ, but also an inconsistent one. This, of course, does not prevent the use of ρ_s in estimating ρ.

The table which follows is a table of $E(\rho_s)$ as given by (2) for $n = 5, 10, 20$ and ∞, $\rho = 0(0\cdot1)\,0\cdot9$. The lower part of the table, for comparison, shows the expected values of r, the sample product-moment correlation coefficient, for the same values. These were taken from the tables in Soper, etc. (1917). Interpolation for $E(\rho_s)$ is linear in $(n+1)^{-1}$ and nearly linear in n^{-1}.

$\rho =$	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9
$E(\rho_s)$ $n = 5$	0·0797	0·1597	0·2408	0·3233	0·4080	0·4958	0·5883	0·6881	0·8022
$n = 10$	0·0869	0·1741	0·2620	0·3511	0·4419	0·5349	0·6313	0·7326	0·8428
$n = 20$	0·0910	0·1823	0·2742	0·3671	0·4613	0·5573	0·6559	0·7580	0·8659
$n = \infty$	0·0955	0·1913	0·2876	0·3849	0·4826	0·5819	0·6829	0·7859	0·8915
$E(r)$ $n = 5$	0·0884	0·1773	0·2671	0·3584	0·4517	0·5480	0·6482	0·7541	0·8687
$n = 10$	0·0946	0·1896	0·2850	0·3813	0·4787	0·5776	0·6785	0·7819	0·8887
$n = 20$	0·0974	0·1950	0·2928	0·3911	0·4900	0·5896	0·6902	0·7919	0·8951

Before considering the calculation of $\mathrm{var}(\rho_s)$, which we can find from $\mathrm{var}(s)$ and so from $E(s^2)$, we must consider the problem of evaluating the total probability in the positive part of a quadrivariate normal distribution.

Suppose that x_1, x_2, x_3, x_4 are distributed in a quadrivariate normal distribution with correlation matrix

$$\begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} \\ & 1 & \rho_{23} & \rho_{24} \\ & & 1 & \rho_{34} \\ & & & 1 \end{pmatrix}.$$

We write

$$T\begin{pmatrix} \rho_{12} & \rho_{13} & \rho_{14} \\ & \rho_{23} & \rho_{24} \\ & & \rho_{34} \end{pmatrix} = pr\{x_1 > 0, x_2 > 0, x_3 > 0, x_4 > 0\}.$$

This will be independent of the variances of the x's which we suppose all equal to unity. Then the characteristic function of the distribution is

$$\phi(t_1, t_2, t_3, t_4) = \exp\{-\tfrac{1}{2}t_1^2 - \tfrac{1}{2}t_2^2 - \tfrac{1}{2}t_3^2 - \tfrac{1}{2}t_4^2 - \rho_{12}t_1t_2 - \rho_{13}t_1t_3 - \rho_{14}t_1t_4 - \rho_{23}t_2t_3 - \rho_{24}t_2t_4 - \rho_{34}t_3t_4\},$$

and we then have

$$T\begin{pmatrix} \rho_{12} & \rho_{13} & \rho_{14} \\ & \rho_{23} & \rho_{24} \\ & & \rho_{34} \end{pmatrix}$$

$$= \frac{1}{16\pi^4}\int_0^\infty\int_0^\infty\int_0^\infty\int_0^\infty dx_1\,dx_2\,dx_3\,dx_4 \int_{-\infty}^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty \phi(t_1, t_2, t_3, t_4)\exp\{-i\sum_1^4 t_i x_i\}\,dt_1\,dt_2\,dt_3\,dt_4.$$

Now

$$\phi(t_1, t_2, t_3, t_4)$$

$$= \exp\left\{-\tfrac{1}{2}\sum_1^4 t_i^2\right\}\sum_{l=0}^\infty\sum_{m=0}^\infty\sum_{n=0}^\infty\sum_{p=0}^\infty\sum_{q=0}^\infty\sum_{r=0}^\infty \frac{(-1)^{l+m+n+p+q+r}\rho_{12}^l\rho_{13}^m\rho_{14}^n\rho_{23}^p\rho_{24}^q\rho_{34}^r}{l!\,m!\,n!\,p!\,q!\,r!}$$

$$\times\, t_1^{l+m+n}t_2^{l+p+q}t_3^{m+p+r}t_4^{n+q+r},$$

and we can therefore write

$$T\begin{pmatrix} \rho_{12} & \rho_{13} & \rho_{14} \\ & \rho_{23} & \rho_{24} \\ & & \rho_{34} \end{pmatrix} = \sum_{l=0}^\infty\sum_{m=0}^\infty\sum_{n=0}^\infty\sum_{p=0}^\infty\sum_{q=0}^\infty\sum_{r=0}^\infty A_{lmnpqr}\rho_{12}^l\rho_{13}^m\rho_{14}^n\rho_{23}^p\rho_{24}^q\rho_{34}^r,$$

where

$$A_{lmnpqr} = \frac{(-1)^{l+m+n+p+q+r}G_{l+m+n}G_{l+p+q}G_{m+p+r}G_{n+q+r}}{l!\,m!\,n!\,p!\,q!\,r!},$$

and

$$G_s = \frac{1}{2\pi}\int_0^\infty dx\int_{-\infty}^\infty t^s\exp\{-\tfrac{1}{2}t^2 - itx\}\,dt.$$

Now
$$\frac{1}{2\pi}\int_{-\infty}^{\infty} t^s \exp\{-\tfrac{1}{2}t^2 - itx\}\,dt = \frac{1}{2\pi(-i)^s}\left(\frac{d}{dx}\right)^s \int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}t^2 - itx\}\,dt$$
$$= \frac{1}{(-i)^s}\left(\frac{d}{dx}\right)^s \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}x^2}.$$

From this it follows that for $s = 0$, $G_s = \frac{1}{2}$. Consider $s > 0$. Then

$$G_s = \int_0^{\infty} \frac{1}{(-i)^s(2\pi)^{\frac{1}{2}}}\left(\frac{d}{dx}\right)^s e^{-\frac{1}{2}x^2}\,dx.$$
$$= \frac{-1}{(-i)^s(2\pi)^{\frac{1}{2}}}\left[\left(\frac{d}{dx}\right)^{s-1} e^{-\frac{1}{2}x^2}\right]_{x=0}$$
$$= 0 \qquad \text{when } s \text{ is even}$$
$$= \frac{(2m)!}{i(2\pi)^{\frac{1}{2}}\, 2^m m!} \quad \text{when } s \text{ is odd and equal to } 2m+1, \text{ where } m = 0, 1, 2, \ldots.$$

Our final result is therefore

$$T\begin{pmatrix} \rho_{12} & \rho_{13} & \rho_{14} \\ & \rho_{23} & \rho_{24} \\ & & \rho_{34} \end{pmatrix} = \sum_{l=0}^{\infty}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\sum_{r=0}^{\infty} \frac{(-1)^{l+m+n+p+q+r} G_{l+m+n} G_{l+p+q} G_{m+p+r} G_{n+q+r}}{l!\,m!\,n!\,p!\,q!\,r!} \rho_{12}^l \rho_{13}^m \rho_{14}^n \rho_{23}^p \rho_{24}^q \rho_{34}^r$$
$$= \frac{1}{16} + \frac{1}{8\pi}(\rho_{12} + \rho_{13} + \rho_{14} + \rho_{23} + \rho_{24} + \rho_{34})$$
$$+ \frac{1}{4\pi^2}(\rho_{12}\rho_{34} + \rho_{13}\rho_{24} + \rho_{14}\rho_{23}) + \ldots, \tag{3}$$

but not all the terms are positive. For example, the coefficient of

$$\rho_{12}\rho_{13}\rho_{14}(l = m = n = 1,\ p = q = r = 0) \quad \text{is} \quad -\frac{1}{4\pi^2}.$$

Expressions similar to the above have also been given by M. G. Kendall (1941, 1945) and have also been given in the lecture courses of Prof. A. C. Aitken as is stated in Kendall (1941). Kendall's version of the formula for four variables needs to be interpreted in the light of an earlier comment that the leading term is conventionally determined and in any case omits a power of -1.

We now consider the problem of finding $\operatorname{var}(\rho_s)$. Since

$$\operatorname{var}(\rho_s) = \frac{144}{n^2(n^2-1)^2}\operatorname{var}(s)$$

it will be sufficient to find $\operatorname{var}(s) = E(s^2) - [E(s)]^2$. Now

$$s^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{r=1}^{n} H(x_i - x_j)\,H(y_i - y_k)\,H(x_p - x_q)\,H(y_p - y_r). \tag{4}$$

Terms in this expansion for which $i = j$, $i = k$, $p = q$ or $p = r$ are zero. If we write

$$X_1 = x_i - x_j, \quad X_2 = y_i - y_k, \quad X_3 = x_p - x_q, \quad X_4 = y_p - y_r,$$

the quantities X_1, X_2, X_3, X_4 will be distributed in a quadrivariate normal distribution and the chance that they are all positive will be

$$T\begin{pmatrix} \frac{1}{2}\rho(1+\delta_{jk}) & \frac{1}{2}(\delta_{ip} - \delta_{iq} - \delta_{jp} + \delta_{jq}) & \frac{1}{2}\rho(\delta_{ip} - \delta_{ir} - \delta_{jp} + \delta_{jr}) \\ & \frac{1}{2}\rho(\delta_{ip} - \delta_{iq} - \delta_{kp} + \delta_{kq}) & \frac{1}{2}(\delta_{ip} - \delta_{ir} - \delta_{kp} + \delta_{kr}) \\ & & \frac{1}{2}\rho(1+\delta_{qr}) \end{pmatrix}, \tag{5}$$

where δ_{ij} is the Kronecker δ equal to unity if $i = j$ and zero otherwise. Using (3) we can now evaluate

$$E\{H(x_i - x_j)\,H(y_i - y_k)\,H(x_p - x_q)\,H(y_p - y_r)\} = pr\{X_1 > 0, X_2 > 0, X_3 > 0, X_4 > 0\},$$

and we must calculate (5), using (3), for each type of set of suffixes in the terms of (4). The number of times each such type of term occurs in (4) will be $n(n-1)\ldots(n-s+1)$, where s is the number of distinct suffixes. From a theoretical point of view this solves the problem but in practice, since there are a fairly large number of different terms which arise from the various ways of identifying suffixes in (4), it is probably only barely possible to calculate $\operatorname{var}(\rho_s)$ for any given value of ρ. It is certainly not practical as a routine.

The above investigation may be compared with the similar investigations on the sampling distribution of Kendall's τ (Greiner, 1909; Esscher, 1924; Kendall, 1948) and the relative simplicity of the latter, especially in the formulae for the variance, brings out once again the superiority of τ as a measure of rank correlation.

REFERENCES

ESSCHER, F. (1924). On a method of determining correlation from the ranks of variables. *Skand. AktuarTidskr.* **7**, 201.

GREINER, R. (1909). Über das Fehlersystem der Kollektiv-masslehre. *Z. Math. Phys.* **57**, 225.

HÖFFDING, W. (1948). A class of statistics with asymptotically normal distributions. *Ann. Math. Statist.* (In the Press.)

KENDALL, M. G. (1941). Proof of relations connected with the tetrachoric series and its generalization. *Biometrika*, **32**, 196.

KENDALL, M. G. (1943). *The Advanced Theory of Statistics*, vol. 1 (3rd ed. 1947). Charles Griffin and Co.

KENDALL, M. G. (1945). On the analysis of oscillatory time-series. *J.R. Statist. Soc.* **108**, 93.

KENDALL, M. G. (1948). *Rank Correlation Methods*. London: Charles Griffin & Co. (In the Press.)

PEARSON, K. (1907). Mathematical contributions to the theory of evolution. XVI. On further methods of determining correlation. *Drap. Co. Mem. biom. Ser.* Cambridge University Press.

SHEPPARD, W. F. (1898). On the application of the theory of error to cases of normal distributions and normal correlations. *Philos. Trans.* A, **192**, 101.

SOPER, H. E. AND OTHERS (1917). On the distribution of the correlation coefficient in small samples. *Biometrika*, **11**, 328.

Tests of significance in the variate difference method

BY N. L. JOHNSON

1. The variate difference method has been discussed by many writers, and in particular by Yule (1921), Anderson (1923, 1926, 1927) and Tintner (1940). It is a method of isolating the random part of certain types of time series and is briefly described in § 2 below.

2. Let $u_1, u_2, \ldots, u_n$ be successive observations at equal intervals of time, forming a time series. Suppose that it is reasonable to assume that

$$u_t = f_t + z_t, \tag{1}$$

where (i) f_t is a 'smooth trend' such that for some integer K (and hence for all larger integers)

$$\Delta^K f_t = 0;$$

(ii) z_t is a 'random residual' with expected value zero and standard deviation σ, independent of t;
(iii) $z_1, z_2, \ldots, z_n$ are mutually independent.

Under these conditions, if the series (u_t) be differenced K times, the f_t terms will be eliminated and, in fact,

$$\Delta^K u_t = \Delta^K z_t. \tag{2}$$

Furthermore $\Delta^{K+1}u_t = \Delta^{K+1}z_t$, $\Delta^{K+2}u_t = \Delta^{K+2}z_t$ and so on.

Since these relations hold, we have, for $k \geqslant K$,

$$E(\Delta^k u_t) = 0, \tag{3·1}$$

$$E[\{\Delta^k u_t\}^2] = {}^{2k}C_k\sigma^2. \tag{3·2}$$

Thus, if $k \geqslant K$

$$S_k = [(n-k)\,{}^{2k}C_k]^{-1} \sum_{t=1}^{n-k} \{\Delta^k u_t\}^2 \tag{4}$$

is an unbiased estimate of σ^2.

If $\Delta f_t = 0$, i.e. if the original series is random

$$S_0 = (n-1)^{-1} \sum_{t=1}^{n} (u_t - \bar{u})^2 \tag{5}$$

is an unbiased estimate of σ^2.

The sequence of statistics $S_0, S_1, S_2 \ldots$ is computed from the observed values $u_1, u_2, \ldots, u_n$; S_k being defined by (4) for $k \geqslant 1$ and by (5) for $k = 0$. This sequence may be used as an indicator of the order of difference at which the non-random elements are first eliminated. Generally the original series is not

random and so S_1 is usually considerably smaller than S_0. As k increases the ratio S_{k+1}/S_k should approach unity. The value of k at and after which S_{k+1}/S_k stays sufficiently close to unity is taken as the order of difference necessary to eliminate the non-random terms f_i.

3. In order to test precisely whether the ratio S_{k+1}/S_k is in fact 'sufficiently close' to unity it is necessary to assume a form of distribution for the z_i's. It is supposed that the assumption of a normal form of distribution for the z_i's will not give rise to serious error. Even with this assumption it is very difficult to obtain exact significance levels for the ratio S_{k+1}/S_k, owing to the existence of correlations of various degrees of intensity between the differences involved.

Tintner suggested a method of overcoming this drawback. He proposed that certain sets of the kth and $(k+1)$th differences should be selected in such a way that any selected difference of either order should be uncorrelated with any other selected difference of either order. The sets to be selected were of the type:

$$(\Delta^k u_r) \quad r = t, \quad t+(2k+3), \quad t+2(2k+3), \quad \dots, \quad t+(j-1)(2k+3);$$

$$(\Delta^{k+1} u_s) \quad s = t+k+1, \quad t+k+1+(2k+3), \quad \dots, \quad t+k+1+(j-1)(2k+3).$$

t may have any integral value from 1 to $(2k+3)$ inclusive, each value giving rise to a different selection; j has the largest value possible. In any one selection none of the quantities $\Delta^k u_r$, $\Delta^{k+1} u_s$ have a u_t in common, and their independence is thereby assured. If the non-random elements have been removed by taking kth differences, then $(\Delta^k u_r)$, $(\Delta^{k+1} u_s)$ should be sequences of independent normal variables, each with expected value zero and with variances in the ratio

$$^{2k}C_k / ^{2k+2}C_{k+1} = \tfrac{1}{2}(k+1)/(2k+1).$$

Tintner suggests using as test criterion

$$F = \frac{\sum_r \{\Delta^k u_r\}^2}{\sum_s \{\Delta^{k+1} u_s\}^2} \frac{2(2k+1)}{k+1} \quad \text{or} \quad z = \tfrac{1}{2}\log_e F. \tag{6}$$

Standard tables of F or z, entered with degrees of freedom j, j would then provide exact significance limits for the suggested criteria.

4. Although this method of selection gives a test which is exact, provided the assumptions upon which it is based are correct, it involves the sacrifice of a considerable proportion of the data available. Thus only about one out of each $(2k+3)$ members of the two difference columns under comparison is used explicitly in the test criteria. This is necessary if the correlation between *any* two selected differences is to be zero.

It is possible, however, to make use of somewhat more of the data available, still subjecting the results to a fairly simple exact test, analogous to that proposed by Tintner. Consider, in fact, the modified method of selection leading to the sequence of pairs of differences

$$(\Delta^k u_r, \Delta^{k+1} u_r) \quad r = t, \quad t+k+2, \quad t+2(k+2)\dots, \quad t+(j'-1)(k+2). \tag{7}$$

Here t may be any integer from 1 to $(k+2)$ inclusive, and j' has its largest possible value.

If non-random elements have been eliminated in the kth differences then

(i) the correlation between corresponding differences $\Delta^k u_r$ and $\Delta^{k+1} u_r$ is $-[\tfrac{1}{2}(2k+1)/(k+1)]^{\frac{1}{2}}$;

(ii) any difference of one order is correlated only with the corresponding difference of the other order;

(iii) the expected values and ratio of variances of the sequences $(\Delta^k u_r)$, $(\Delta^{k+1} u_r)$ are the same as in Tintner's method.

An exact test based on the selection (7) would utilize about twice as many differences as would Tintner's test for the same time series. For example, we have the following comparison between Tintner's method of selection and method (7) in the special case $t = 1, k = 2$.

	Tintner's Method		Present Method (7)	
	$\Delta^2 u_r$	$\Delta^3 u_s$	$\Delta^2 u_r$	$\Delta^3 u_r$
	$\Delta^2 u_1 = u_3 - 2u_2 + u_1$	$\Delta^3 u_4 = u_7 - 3u_6 + 3u_5 - u_4$	$\Delta^2 u_1 = u_3 - 2u_2 + u_1$	$\Delta^3 u_1 = u_4 - 3u_3 + 3u_2 - u_1$
	$\Delta^2 u_8 = u_{10} - 2u_9 + u_8$	$\Delta^3 u_{11} = u_{14} - 3u_{13} + 3u_{12} - u_{11}$	$\Delta^2 u_5 = u_7 - 2u_6 + u_5$	$\Delta^3 u_5 = u_8 - 3u_7 + 3u_6 - u_5$
	etc.	etc.	etc.	etc.
Variance	$6\sigma^2$	$20\sigma^2$	$6\sigma^2$	$20\sigma^2$
Correlation	0		$-\sqrt{(5/6)}$	

The gain in efficient use of the data would not, however, be as great as this would suggest, since the correlation is high between the differences already in Tintner's selection and the additional differences introduced in (7). Nevertheless, a certain amount of extra information would be used in a test based on (7). Such an exact test, analogous to Tintner's test, is derived in § 5.

5. In the case of Tintner's method of selection, the hypothesis which should be tested is that the expected values of each of the differences is zero. The ratio of the variances of the differences in the two sequences is known. The alternative hypotheses specify non-zero values of the expected values, the ratio of the variances remaining unchanged. It is not necessary that these non-zero expected values be constant for either of the sequences of differences involved. Tintner's criterion is, however, appropriate to the case where the alternative hypotheses specify different values for the ratio of the variances, the expected values remaining at zero (i.e. the hypothesis tested is that the variances are in a certain ratio, it being assumed that the expected values are zero). This system of hypotheses seems to be a reasonable approximation to the situation, as may be appreciated by regarding the differences $\Delta^k f_r, \Delta^{k+1} f_r$, as random variables.

A similar approach in the case of selection (7) leads to the conclusion that the hypothesis to be tested may be stated as follows: 'The variances of the two sequences of differences are in a certain ratio, and the correlation between corresponding differences has a certain value. It is assumed that all expected values are zero.' The alternative hypotheses specify other values for the ratio of the variances and the correlation, all expected values being supposed to remain at zero.

A test appropriate to this situation may be obtained by a slight modification of a result due to Hsu (1940), who used Neyman and Pearson's likelihood ratio method to derive a number of tests of hypotheses regarding two normally correlated variables. The hypothesis to be tested is, in fact, nearly the same as Hsu's hypothesis H_4 and the appropriate test criterion is similar to his criterion L_4. This similarity is not, however, apparent when the special symbols and values of the present problem are inserted, giving a test criterion

$$L = \frac{T_{k,k} T_{k+1,k+1} - T^2_{k,k+1}}{(2k+1)[T_{k,k} + \frac{1}{2}(k+1)(2k+1)^{-1} T_{k+1,k+1} + T_{k,k+1}]^2}, \tag{8}$$

where $T_{p,q} = \sum_r \Delta^p u_r \Delta^q u_r$.

L must lie between 0 and 1. Low values of L are regarded as significant of departure from the hypothesis tested, indicating that the 'non-random' terms have not been eliminated in the kth differences. If non-random variations are in fact absent from the kth differences, the probability density function of L is

$$p(L) = \tfrac{1}{2}(j'-1) L^{\frac{1}{2}(j'-3)} \quad (0 \leq L \leq 1). \tag{9}$$

It will be recalled that j' is the number of differences in each of the sequences $(\Delta^k u_r), (\Delta^{k+1} u_r)$. From (9) it follows that L_α, the significance limit corresponding to a probability α of rejecting the hypothesis when it is valid, satisfies the equation

$$L_\alpha^{\frac{1}{2}(j'-1)} = \alpha,$$

i.e.

$$\log_{10} L_\alpha = (2 \log_{10} \alpha)/(j'-1). \tag{10}$$

If $\log_{10} L$ be taken as test criterion, the significance limits may be calculated very easily by means of (10). We note that the 5 % limit for $\log_{10} L$ is $-2{\cdot}6/(j'-1)$, and the 1 % limit is $-4{\cdot}0/(j'-1)$.

6. We shall now apply selection (7) to the series of American wheat-flour prices for the years 1890–1937, which is analysed by Tintner. There are forty-eight observations in this series. Taking $t = 1$, and comparing the first and second order differences (i.e. $k = 1$) we have the sequences

$$\Delta u_1, \quad \Delta u_4, \quad \Delta u_7, \quad \ldots, \quad \Delta u_{46};$$
$$\Delta^2 u_1, \quad \Delta^2 u_4, \quad \Delta^2 u_7, \quad \ldots, \quad \Delta^2 u_{46};$$

so that $j' = 16$. (It may be noted that the sequences of differences in Tintner's selection 1-A each contain 9 members.) The sums of squares and products are

$$T_{1,1} = 31{\cdot}259830, \quad T_{1,2} = -23{\cdot}842300, \quad T_{2,2} = 29{\cdot}137173,$$

whence

$$L = 0{\cdot}3889,$$

and

$$\log_{10} L = -0{\cdot}4102.$$

For $j' = 16$, the 1 % limit for $\log_{10} L$ is $-0{\cdot}27$ and the 5 % limit is $-0{\cdot}17$. The calculated value of $\log_{10} L$ is less than the 1 % limit, and it seems unlikely that the non-random terms have been eliminated in the first differences.

Now taking $k = 2$ (and $t = 1$ as before) we have the sequences of differences

$$\Delta^2 u_1, \quad \Delta^2 u_5, \quad \Delta^2 u_9, \quad \ldots, \quad \Delta^2 u_{45}; \qquad \Delta^3 u_1, \quad \Delta^3 u_5, \quad \Delta^3 u_9, \quad \ldots, \quad \Delta^3 u_{45};$$

so that in this case $j' = 12$. The sums of squares and products are

$$T_{2,2} = 12{\cdot}030127; \quad T_{2,3} = -10{\cdot}330714; \quad T_{3,3} = 53{\cdot}809942$$

whence

$$L = 0{\cdot}3396,$$

and

$$\log_{10} L = -0{\cdot}4690.$$

For $j' = 12$, the 1 % limit for $\log_{10} L$ is -0.36 and the 5 % limit is -0.24. Again it seems unlikely that non-random terms have been eliminated.

Comparing third and fourth order differences ($k = 3$) we have, using the selection corresponding to $t = 1$

$$T_{3,3} = 102{\cdot}200898, \quad T_{3,4} = -188{\cdot}615311, \quad T_{4,4} = 358{\cdot}329689,$$

giving $L = 0{\cdot}5862$ and $\log_{10} L = -0{\cdot}2320$. The appropriate 1 % limit ($j' = 9$) is $-0{\cdot}50$ and the 5 % limit is $-0{\cdot}33$. This test supports the hypothesis that non-random terms have been eliminated in the third order differences. Before a final decision is reached, of course, further tests would be made, using other possible values of t, and dealing with higher orders of differences. The above calculations should, however, be sufficient to make clear the mode of application of the method of selection (7) and its associated test of significance.

7. It is not essential that the sequence of pairs of differences be chosen as in (7) above. Any sequence satisfying conditions (ii) and (iii) of § 4 may be used, and a corresponding criterion, similar to (8), obtained by inserting the appropriate terms in the $T_{p,q}$'s and using the correct value of the correlation coefficient between corresponding differences. The sequence

$$(\Delta^k u_{r+1}, \Delta^k u_r) \quad r = t, \quad t+k+2, \quad t+2(k+2), \quad \ldots, \quad t+(j'-1)(k+2)$$

is very similar to (7) and leads, in fact, to identically the same criterion.

Other alternative methods of selection may aim at reducing the correlation between corresponding differences. For example, in the sequence

$$(\Delta^k u_r, \Delta^{k+1} u_{r+p})\,(p \leqslant k) \quad r = t, \quad t+k+2+p, \quad t+2(k+2+p), \quad \ldots,$$

this correlation is

$$(-1)^{p+1}\,{}^{2k+1}C_{p+k+1}/({}^{2k}C_k\,{}^{2(k+1)}C_{k+1})^{\frac{1}{2}} = (-1)^{p+1}\frac{k(k-1)\ldots(k-p+1)}{(k+2)(k+3)\ldots(k+p+1)}\left(\frac{2k+1}{2k+2}\right)^{\frac{1}{2}}.$$

Unfortunately the more the correlation is reduced the greater the proportion of data eliminated by this method of selection.

A further possible method, possessing the advantage of symmetry, could be based on the comparison of kth and $(k+2)$th order differences, using the sequence

$$(\Delta^k u_{r+1}, \Delta^{k+2} u_r) \quad r = t, \quad t+k+3, \quad t+2(k+3), \ldots,$$

in which the correlation between corresponding differences is

$$-[(2k+1)(k+2)/(2k+3)(k+1)]^{\frac{1}{2}}.$$

REFERENCES

ANDERSON, O. (1923). *Biometrika*, **15**, 134.

ANDERSON, O. (1926). *Biometrika*, **18**, 293.

ANDERSON, O. (1927). *Biometrika*, **19**, 53.

HSU, C. T. (1940). *Ann. Math. Statist.* **11**, 410.

TINTNER, G. (1940). *The Variate Difference Method.* Cowles Commission, Principia Press, Bloomington, Ind.

YULE, G. U. (1921). *J. Roy. Statist. Soc.* **84**, 497.

REVIEW

The Advanced Theory of Statistics. Vol. II, pp. 1–521. By M. G. KENDALL, M.A. London: Griffin and Co., Ltd. 1946. Price: 50*s*.

There is only a small measure of agreement among writers on the theory of statistics as to what should constitute an advanced course in the subject. There is no common view about either the nature of the problems which should be discussed or the methods by which they should be solved. Disagreement about the actual details of solutions of theoretical problems is, perhaps, only a passing phase and its importance may be exaggerated. The way in which the problems themselves are posed, however, reveals more fundamental cleavages of opinion which may not be so easily reconciled.

An important element which renders difficult the formulation of theoretical problems is the uncertain relation existing between the theory and the practice of the application of statistical methods, as soon as we advance beyond the most elementary stages. In discussing this relation it is convenient to distinguish between the theoretician or, as he is now categorized, the mathematical statistician, on the one hand, and the practical statistician, using statistical methods in some particular field of inquiry, on the other. This distinction may have some validity if it is not pushed too hard. It is true that at one extreme there is a body of workers whose interest in the theory of statistics is primarily as a source of mathematical problems. And, at the other extreme, there are experimentalists who apply statistical methods by rule of thumb, without much concern for the reasoning which is needed to justify them. But, although the modern tendencies in the organization of all the sciences may be such as to force workers into extreme positions and label them accordingly, it must be recognized that, in statistics, the theoretical and the practical investigator are still, fortunately, often one and the same person.

This fact undoubtedly makes for a closer integration of theoretical developments with practical usages, but not, perhaps, to the extent that one might imagine. It is in the nature of theoretical work—almost in its definition—that, whoever may be responsible for it, it should assert the right to an independent existence of its own and refuse to be tied down too closely by considerations of its ultimate usefulness. Even were statistical theory to be cultivated exclusively by writers who at the same time were outstandingly competent in dealing with the more commonplace, everyday, statistical investigations, it would still not escape the tendency to become a purely abstract discipline. The time spent on it would still have to be justified by secondary considerations, such as its aesthetic appeal, educational value and the like, as well as by its ability to provide novel statistical procedures of practical value in the experimental and social sciences.

The tendency of statistical theory towards exclusive concern with abstract generalizations is helped, of course, by those mathematicians who are interested in nothing else. There is, however, as one might expect, a strong reaction against it from statisticians who still maintain a wider outlook. This reaction expresses itself, sometimes explosively, in the form of exasperated criticism of this or that piece of theoretical work on the grounds, either of its irrelevance, or worse, of the misleading impressions which it may give of the true objects of practical statistical investigations. The criticism is not always fair or well directed. Its authors are seldom immune from the failings, real or imaginary, which they castigate in others. But it is sufficient to show that statisticians as a body, however pleased they may be that their subject is gaining increased academic recognition, are yet unwilling to sanction the appearance of two separate subjects—pure and applied statistics—which can develop without much reference one to the other. Practical interests are still able to make themselves felt even in academic developments. Further, since the practical interests of statisticians are so diverse, it is not surprising that a parallel difficulty is found in reaching agreement as to what is the proper subject-matter for the advanced theory of statistics.

Mr M. G. Kendall, in the notable work of which the second volume is here under review, does not attempt to confine the subject within any very strict limits, nor to press upon the reader any very decided views as to what should constitute its most important features. His method is rather to let the truth appear—if appear it can—by making an extensive and detailed survey of the whole range covered by modern contributions.

The first four chapters of this second volume are concerned with the fundamental problem of estimating population parameters from sample values and of assigning limits to such parameters on a probability basis. This has in the past produced a most controversial literature, revolving round the question of the applicability or otherwise of the theory of inverse probability and of the various alternatives to this theory which have been proposed. Mr Kendall devotes most attention to the approach which does not make explicit use of the concept of inverse probability. Some readers might

consider that he should have given more space to the alternative position which is founded on Bayes's Theorem and its corollaries. However, in a sense, Harold Jeffreys's excellent restatement of this position in his *Theory of Probability* renders a lengthy description of it unnecessary, unless, indeed, Mr Kendall's object had been to make a critical comparison and evaluation of the inverse and direct probability approaches. Mr Kendall is content to describe rather than to judge. One might, perhaps, criticize him in this section of the book for carrying his non-committal attitude too far in the face of the irreconcilable viewpoints which emerge from the discussion. He could, possibly, have pressed his own ideas upon us more strongly in certain passages, without departing from the high standard of fairness which he maintains throughout in representing the ideas of his statistical colleagues.

The problem of scientific inference discussed in these first four chapters, although a fertile source of argument, is, strangely enough, one which, in whatever way it is settled, does not appear to influence much the actual way in which scientific investigation is carried out in practice. The next five chapters are concerned with a miscellaneous range of questions, which, while not having the same attraction for the pure theorist, are more closely related to those statistical methods used most frequently in the everyday interpretation of experimental data. They include, for instance, some discussion of the theory of the analysis of variance and of the role which randomization plays in ensuring the applicability of this theory in practice. The attitude which one adopts towards the theoretical questions discussed in this part of the book can affect quite directly the choice between the alternative experimental procedures which may be at one's disposal in some specific inquiry.

The succeeding two chapters revert to a more generalized treatment following the same lines as the earlier ones. They are again concerned with the fundamental problem of scientific inference, although now from the restricted angle of significance testing. They furnish another illustration of the author's remarkable ability to reproduce the spirit as well as the substance of the original contributions covered by his survey. They again suffer, perhaps, from some lack of critical evaluation on the author's own part.

The next chapter deals with multivariate analysis, including discriminant functions. From the theoretical standpoint this is a straightforward development from the univariate and bivariate cases. Some new problems are raised, but much is simply generalization. Of the possible applications of these recent developments it is difficult to speak confidently. It is clear that in many fields multivariate analysis of the type here described can only proceed on the basis of many dubious assumptions as to the nature of possible connexions between the variates. Moreover, simple interpretation of the results of a multivariate analysis in terms which convey much to the layman is difficult. On the other hand, in certain cases where a very large number of variates are measured (e.g. in intelligence testing) some logical way of dealing with the results is needed, and one may in some instances be sure enough of one's assumptions to exploit this recent theoretical work.

The two final chapters deal with Time Series, a subject perhaps most studied by economists, but not one which has borne much fruit. Trade cycles have been sufficiently talked about, but there has been little analysis of data demonstrating their existence in an unequivocal fashion. Mr Kendall is the foremost representative of a school of thought which holds that the search for regular periodicities in economic data has been largely a waste of time, and that much more is to be hoped for by considering so-called autoregressive schemes, which allow for irregularities in the lengths of periods—albeit irregularities governed by some simple law. It is too soon yet to say whether the new conceptions will be much more successful than the old, but whatever success is obtained will be largely due to Mr Kendall's own efforts to clarify the subject.

Looking back after reading through the two volumes of Mr Kendall's ambitious enterprise, one is forced to the realization that here we have one of the most remarkable compilations that has ever been attempted by a single writer in any branch of science. If several writers had banded together to produce it one would still be impressed by the wide range of topics and viewpoints which it embraces. One might almost be pardoned, indeed, for thinking that 'M. G. Kendall' was a pseudonym standing for the collaboration of several persons. This would, at all events, be a pleasant theory to explain the many excellences of the work, particularly its freedom from personal animosities and the generosity of its references to the contributions of so many different writers. It would also explain the somewhat loose-knit nature of the work judged as a whole. Occasionally one feels the lack of a sufficiently strong common thread running through it and holding it together; such as one would perhaps obtain if the author were inclined to be more trenchant in his criticisms of what other people write and more self-assertive in putting forward his own contributions. But in reply to all possible criticisms, Mr Kendall has the final word, when, in his concluding section, he says: 'Much remains to be done; and this book will have served its purpose if the reader is left with the desire to do some of it himself.' Few authors could by implication have exposed so clearly just what it is that remains to be done.

B.L.W.

SOME FURTHER NOTES ON THE USE OF MATRICES IN POPULATION MATHEMATICS

BY P. H. LESLIE

Bureau of Animal Population, Department of Zoological Field Studies, Oxford University

CONTENTS

1. INTRODUCTION

The use of matrices in population mathematics has been discussed in a previous paper (Leslie, 1945), and some of the properties of the basic matrix representing a system of age-specific fertility and mortality rates have been described both there, and also in an earlier paper by Lewis (1942).† The purpose of the following notes is to enlarge on a few points left over from the earlier work, and in the later sections to extend the use of matrices and vectors to the case of the logistic type of population growth and to the predator-prey type of relationship between two or more populations.

In order to save a troublesome amount of cross-referring, it may perhaps be a convenience if the definitions and properties of the basic vectors and matrices are summarized here, and also if a brief account is given of the various transformations which are at one time or another used in the theoretical development. For fuller details reference may be made to the appropriate section of the original paper.

As before, for the sake of simplicity, the female population only will be considered, and the same unit of age will be adopted as that of time. If m to $m+1$ is the last age group in the complete life-table distribution defined by $L_x = \int_x^{x+1} l_x dx$ (taking $l_0 = 1$), and we put

P_x $(x = 0, 1, 2, \ldots, m-1) = L_{x+1}/L_x =$ the probability that a female aged x to $x+1$ at time t will be alive in the age group $x+1$ to $x+2$ at time $t+1$,

F_x $(x = 0, 1, 2, \ldots, m) =$ the number of daughters born in the interval t to $t+1$ per female alive aged x to $x+1$ at time t, who will be alive in the age group 0 to 1 at time $t+1$,

† At the time my original paper was published I was not aware that the same problem had already been investigated by Lewis (1942). This author establishes the form of the basic matrix and discusses a number of its properties, including the role of the dominant latent root and the form of the stable age distribution. He suggests that the rapidity with which an arbitrary age distribution settles down to the latter form will depend on the difference between the dominant and subdominant root of the characteristic equation, and he also discusses the type of matrix in which there is only a single non-zero element in the first row. It is clear, therefore, that unwittingly I covered a good deal of ground which had already been covered by him. I am indebted to Prof. M. S. Bartlett and Dr S. Vajda for this reference.

we are led to consider the square matrix M of order $m+1$ which has the F_x figures in the first row and the P_x figures in the subdiagonal immediately below the principal diagonal. For many purposes, however, it may not be necessary to deal with the matrix M as a whole. Thus, if $x = k$ is the last age group within which reproduction occurs, all the F_x figures for $x > k$ will be zero and the determinant $|M| = 0$. Partitioning the matrix symmetrically at this point the principal, non-singular, submatrix is

$$A = \begin{bmatrix} F_0 & F_1 & F_2 & F_3 & \cdots & F_{k-1} & F_k \\ P_0 & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & P_1 & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & P_2 & \cdot & \cdots & \cdot & \cdot \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdots & P_{k-1} & \cdot \end{bmatrix}.$$

As before an arbitrary age distribution will be written as the column vector ξ, different age distributions being distinguished by different subscripts. The number of elements composing a ξ vector may be either $m+1$ or $k+1$ depending on whether the particular age distribution considered is complete, or confined only to the pre-reproductive and reproductive age groups. Associated with each ξ_x there is a uniquely determined vector η_x, which in matrix notation is written as a row vector, the square of the length of the vector ξ_x being given by the scalar product $\eta_x \xi_x$. If the age distribution ξ_x is complete, consisting of $m+1$ elements, the last $m-k$ elements of the associated vector η_x will all be zero. Generally speaking, however, the post-reproductive age groups can be neglected, more particularly in the theoretical development, and unless otherwise stated it will be assumed that we are dealing with η and ξ vectors consisting of $k+1$ elements which are subject to the system of rates represented by the submatrix A.

It was shown in the previous paper (Leslie, 1945, §5) that it is convenient for many purposes to pass to a new frame of reference, the vectors η and ξ and the matrix A undergoing the non-singular linear transformations

$$\eta = \phi H, \quad \xi = H^{-1}\psi, \quad B = HAH^{-1},$$

where H is a diagonal matrix with elements $(P_0 P_1 P_2 \ldots P_{k-1})$, $(P_1 P_2 P_3 \ldots P_{k-1})$, $\ldots$, $(P_{k-2} P_{k-1})$, P_{k-1}, 1, which are derived entirely from the life table. (If the matrix M is the subject of the transformation instead of A, the matrix H may be suitably enlarged and will include all the P_x figures down to P_{m-1}.) It will be noted that in this collineatory transformation the square of the length of a vector is an invariant, and that the matrices A and B have the same characteristic equation and, therefore, the same latent roots.

The effect of this transformation on the elements of A is to replace the P_x figures in the principal subdiagonal by a series of units, and thus to reduce A to the rational canonical form. In biological terms it is equivalent to transforming the original population into one in which all the individuals live until the span of reproductive life is completed at the age of $x = k+1$. This imaginary type of population, with which in many ways it is more convenient to work, might be termed the canonical population.

When the relation between two column vectors is such that

$$B\psi_a = \lambda\psi_a,$$

where λ is a scalar, then ψ_a is termed a stable ψ appropriate to the matrix B. Similarly in the case of initial row vectors, if

$$\phi_a B = \lambda \phi_a,$$

then ϕ_a is a stable ϕ appropriate to B.

It may be shown that corresponding to each distinct latent root λ_a of the characteristic equation of B,

$$|B - \lambda I| = 0,$$

there is a pair of stable vectors ϕ_a and ψ_a which in the usual way may be normalized so that $\phi_a \psi_a = 1$. In the case when all the $k+1$ latent roots of B are distinct, the normalized stable ψ form a set of $k+1$ independent and mutually orthogonal vectors of unit length, and any arbitrary ψ_x may be expanded in terms of them, viz.

$$\psi_x = c_1 \psi_1 + c_2 \psi_2 + c_3 \psi_3 + \ldots + c_{k+1} \psi_{k+1},$$

where the coefficients c_a may be either real or complex. Similarly the associated row vector ϕ_x can be expanded in terms of the stable ϕ,

$$\phi_x = \bar{c}_1 \phi_1 + \bar{c}_2 \phi_2 + \ldots + \bar{c}_{k+1} \phi_{k+1},$$

where $\bar{c}_a$ is the complex conjugate of c_a in the expansion of ψ_x. Similarly, by transforming back to the original co-ordinate system, any arbitrary ξ_x can be expanded in terms of the stable ξ and its associated vector η_x in terms of the stable η.

Since only one of the latent roots, and this the dominant one of the matrix B, is real and positive, only one of the stable ψ will consist of real and positive elements. It is this stable $\xi_1 = H^{-1}\psi_1$, associated with the dominant root λ_1, which is ordinarily referred to as the stable age distribution appropriate to a given set of age-specific fertility and mortality rates. The relation between the inherent rate of increase (r) and the dominant root of the matrix is given by

$$\log_e \lambda_1 = r.$$

There is one further transformation of the matrix B which is of some theoretical importance. The expansion of an arbitrary ψ_x in terms of the normalized stable ψ may be written in matrix notation as

$$\psi_x = Q c_x,$$

where the columns of the matrix Q consist of the stable ψ arranged from left to right in descending order of the moduli of the roots with which they are associated. In the same way the expansion of an arbitrary ϕ_x may be written

$$\phi_x = \bar{c}'_x U,$$

where $\bar{c}'_x$ is the transposed complex conjugate of the vector c_x, and the rows of the matrix U are formed by the stable ϕ arranged in a similar order from above down. Since the normalized stable vectors have the properties

$$\phi_a \psi_b \begin{cases} = 1 & (a = b), \\ = 0 & (a \neq b), \end{cases}$$

it follows that U and Q are reciprocal matrices ($UQ = I$). In this transformation to an orthogonal co-ordinate system the length of a vector remains an invariant and the matrix B becomes

$$UBQ = UHAH^{-1}Q = C,$$

where C is a diagonal matrix whose elements are the latent roots of B (reduction to classical canonical form).

Since an arbitrary age distribution $\psi_x = H\xi_x$ must necessarily consist of real and positive elements, and since $\psi_x = Qc_x, \phi_x = \bar{c}'_x U$, we have $\phi_x = \psi'_x \bar{U}' U = \psi'_x G$, where G is a

symmetrical matrix of real elements. Thus, in terms of the original co-ordinate system, since H is a diagonal matrix unaltered by transposition,

$$\eta_x = \xi_x' HGH.$$

The matrix HGH, or G if the work is being carried out in terms of the canonical population, has the important property of converting a column vector into the associated row vector. The reciprocal relationship is given by

$$\xi_x = H^{-1}G^{-1}H^{-1}\eta_x',$$

where $G^{-1} = Q\bar{Q}'$. For further properties of the metric matrix G see the previous paper (Leslie, 1945, § 11).

It may perhaps be of interest if the actual values of some of these matrices are given for a simple numerical example, which will be used in some of the later sections in order to illustrate certain points. Although this example is purely a mathematical model bearing no relation to any known species, its properties are the same as those which might be observed for a population of living organisms considered in a small number of age groups, and for convenience biological terms will be used throughout in interpreting the results obtained with this matrix. Suppose, then, we have an entirely imaginary population which can be considered in four age groups, and let the life table or stationary age distribution be given by the L_x values forming the column vector $\{0{\cdot}9, 0{\cdot}7, 0{\cdot}5, 0{\cdot}3\}$. Further let the matrix

$$A = \begin{bmatrix} 0 & 45/7 & 18 & 18 \\ 7/9 & 0 & 0 & 0 \\ 0 & 5/7 & 0 & 0 \\ 0 & 0 & 3/5 & 0 \end{bmatrix}. \tag{1·1}$$

Then, since H is the diagonal matrix with elements $h_{11} = P_0P_1P_2$, $h_{22} = P_1P_2$, $h_{33} = P_2$, $h_{44} = 1$, we have

$$H = \begin{bmatrix} 1/3 & 0 & 0 & 0 \\ 0 & 3/7 & 0 & 0 \\ 0 & 0 & 3/5 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$HAH^{-1} = B = \begin{bmatrix} 0 & 5 & 10 & 6 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

The characteristic equation $| B - \lambda I | = 0$ is, when expanded in powers of λ,

$$\lambda^4 - 5\lambda^2 - 10\lambda - 6 = 0;$$

and the latent roots are therefore $\lambda_1 = 3$; $\lambda_2, \lambda_3 = -1 \pm i$; $\lambda_4 = -1$. In the transformation to the classical canonical form $UBQ = C$, the matrix

$$Q = \frac{1}{\sqrt{(68)}} \begin{bmatrix} 27 & 4{\cdot}9985 + 6{\cdot}4019i & 4{\cdot}9985 - 6{\cdot}4019i & \sqrt{(17)}\,i \\ 9 & 0{\cdot}7017 - 5{\cdot}7002i & 0{\cdot}7017 + 5{\cdot}7002i & -\sqrt{(17)}\,i \\ 3 & -3{\cdot}2010 + 2{\cdot}4992i & -3{\cdot}2010 - 2{\cdot}4992i & \sqrt{(17)}\,i \\ 1 & 2{\cdot}8501 + 0{\cdot}3509i & 2{\cdot}8501 - 0{\cdot}3509i & -\sqrt{(17)}\,i \end{bmatrix},$$

$$U = \frac{1}{\sqrt{(68)}} \begin{array}{c} \begin{array}{cccc} \quad 1 \quad & \quad 3 \quad & \quad 4 \quad & \quad 2 \quad \end{array} \\ \begin{bmatrix} 2{\cdot}8501+0{\cdot}3509i & -3{\cdot}2010+2{\cdot}4992i & -13{\cdot}5488-7{\cdot}4545i & -7{\cdot}4977-9{\cdot}6029i \\ 2{\cdot}8501-0{\cdot}3509i & -3{\cdot}2010-2{\cdot}4992i & -13{\cdot}5488+7{\cdot}4545i & -7{\cdot}4977+9{\cdot}6029i \\ -\sqrt{(17)}\,i & \sqrt{(17)}\,i & 4.\sqrt{(17)}\,i & 6.\sqrt{(17)}\,i \end{bmatrix} \end{array},$$

and
$$G = \bar{U}'U = \begin{bmatrix} 0{\cdot}5072 & -0{\cdot}4484 & -2{\cdot}1539 & -2{\cdot}1982 \\ -0{\cdot}4484 & 0{\cdot}8674 & 1{\cdot}9041 & 1{\cdot}5882 \\ -2{\cdot}1539 & 1{\cdot}9041 & 11{\cdot}2688 & 11{\cdot}2109 \\ -2{\cdot}1982 & 1{\cdot}5882 & 11{\cdot}2109 & 13{\cdot}4245 \end{bmatrix},$$

where in each case the elements have been rounded off to the fourth decimal place.

Since the ath row of the matrix U is the stable vector ϕ_a which is associated with the stable ψ_a vector given by the ath column of Q, it is possible to construct readily from their rows and columns the set of four matrices $S_a = \psi_a \phi_a$, which have the properties (Leslie, 1945, § 9)

$$S_a^2 = S_a, \quad S_a S_b = 0 \ (a \neq b), \quad \sum_a S_a = I.$$

If $f(B)$ is a polynomial of the matrix B, we have when the latent roots of the matrix are distinct,

$$f(B) = \sum_{a=1}^{k+1} f(\lambda_a) S_a.$$

Thus
$$B^t = \lambda_1^t S_1 + \lambda_2^t S_2 + \ldots + \lambda_{k+1}^t S_{k+1},$$

so that in the present example, when the matrix B is raised to a high power and λ_1^t is much greater than all the remaining λ_a^t,

$$B^t \propto \begin{bmatrix} 27 & 81 & 108 & 54 \\ 9 & 27 & 36 & 18 \\ 3 & 9 & 12 & 6 \\ 1 & 3 & 4 & 2 \end{bmatrix},$$

and hence, by transforming back to the original co-ordinate system,

$$H^{-1}B^tH = A^t \propto \begin{bmatrix} 81 & 312{\cdot}4283 & 583{\cdot}2 & 486 \\ 21 & 81{\cdot}0000 & 151{\cdot}2 & 126 \\ 5 & 19{\cdot}2857 & 36{\cdot}0 & 30 \\ 1 & 3{\cdot}8571 & 7{\cdot}2 & 6 \end{bmatrix}$$

for large values of t.

2. The stable female birth-rate

Once the dominant latent root of the matrix has been found, there is one comparatively simple way of calculating the stable age distribution. Thus, working in terms of the canonical population and $m+1$ age groups, the stable ψ_1 appropriate to the root λ_1 may be taken proportional to the column vector $\{\lambda_1^m, \lambda_1^{m-1}, \ldots, \lambda_1, 1\}$, and by operating on this vector with the matrix H^{-1}, the stable age distribution ξ_1 can readily be obtained. The method which was used previously for calculating the stable female birth-rate was then to operate on this distribution with the maternal frequency figures† (m_x) and thus determine the total number

† The maternal frequency m_x is the mean number of live daughters born per unit of time to a female aged x to $x+1$. They are the figures tabulated in the usual type of fertility table and are not the same as the F_x figures forming the first row of the matrix.

of female births which might be expected per unit of time (Leslie, 1945, § 16). Although it seems likely that no very great error would be made in employing these methods, both the stable age distribution and the stable birth-rate can be defined rather more formally for the discontinuous case, and the appropriate equations can be derived for calculating them directly when the work is being carried out in terms of discrete age groups.

Consider at time t a stable age distribution $\xi(t)$ appropriate to the dominant latent root λ of the matrix M, and let n_x ($x = 0, 1, 2, \ldots, m$) be the elements of this column vector. Then by the definition of a stable vector

$$\xi(t-x) = \lambda^{-x}\xi(t).$$

If $B(t)$ = the number of daughters born alive in the whole population in the interval of time t to $t+1$, it is easily seen that since n_x are the number of individuals alive aged x to $x+1$,

$$\begin{aligned} \lambda n_0 &= L_0 B(t), \\ \lambda n_1 &= P_0 n_0 \\ &= L_1 B(t-1), \end{aligned}$$

and in general

$$\lambda n_x = L_x B(t-x).$$

If we put

π_x = the proportion of the stable population alive in the age group x to $x+1$,

and $N(t)$ = the total number of individuals alive in the stable population at time t;

$$\pi_x = \frac{L_x B(t-x)}{N(t+1)}.$$

Defining the birth-rate

$$\beta = B(t)/N(t),$$

we have in the case of the stable population,

$$B(t-x) = \beta N(t-x) = \beta N(t)\lambda^{-x},$$

so that

$$\pi_x = \beta L_x \lambda^{-(x+1)}, \tag{2·1}$$

an expression which defines the matrix stable age distribution. From this it follows, since

$$\sum_0^m \pi_x = 1,$$

that

$$\frac{1}{\beta} = \sum_{x=0}^{m} L_x \lambda^{-(x+1)}. \tag{2·2}$$

This argument for the case of discrete age classes is, of course, developed along lines similar to those followed by Lotka (e.g. 1939, p. 16) for the continuous case, where, if c_x is the proportion of the stable population aged between x and $x+dx$ and b the instantaneous birth-rate,

$$c_x = be^{-rx}l_x \quad \text{and} \quad \frac{1}{b} = \int_0^\infty e^{-rx}l_x\,dx. \tag{2·3}$$

The birth-rate β as defined by (2·2) is, however, a different type of birth-rate to that defined by (2·3). It is the total number of births taking place in the interval of time t to $t+1$ expressed per head of population at time t. If $D(t)$ is the number of deaths occurring in the same interval and $\delta = D(t)/N(t)$,

$$N(t+1) = N(t) + B(t) - D(t),$$

and thus, in the case of the stable population,

$$\lambda = 1 + \beta - \delta.$$

In order to express the relationship between β and b, we might consider that in the continuous case the number of births occurring during the interval of time t to $t+1$ will be given by

$$B(t) = bN(t)\int_0^1 e^{r\tau}\,d\tau,$$

whence
$$\beta = \frac{b}{r}(e^r - 1),$$

or, since $\log_e \lambda = r$,
$$b = \frac{\beta \log_e \lambda}{\lambda - 1}. \tag{2·4}$$

As an illustration of the comparative results obtained by applying these equations, we may take the same imaginary population of *Rattus norvegicus* as was used previously as a numerical example (Leslie, 1945). In the appendix to that paper it was shown that for the given system of fertility and mortality rates the value of r, estimated by the more usual methods of computation, was 0·44565, and that $b = 0·51265$, this value of the birth-rate being obtained by the numerical integration of (2·3). When the system of rates was expressed in the form of a matrix of order 21×21, the dominant root was $\lambda_1 = 1·56246$, or $r = 0·44626$, and using equation (2·2) $\beta = 0·64839$, and from (2·4) $b = 0·5144$. The agreement between these estimates of the stable birth-rate is reasonably close and suggests that when we have already calculated the life table age distribution, which is so often the case, equations (2·2) and (2·4) of this section will provide an alternative method of calculating b, which would save a great deal of the tedious labour involved in the numerical integration of (2·3). Although theoretically it is necessary to consider the entire age span of the life table in applying these equations, this was not done in the present instance. In the numerical example given above the value of the rate of increase is so high that the post-reproductive age groups could be neglected without any very great error.

The stable birth-rate and death-rate of the transformed or canonical population (ψ_1 vector) are perhaps only of academic interest. In this connexion, however, there is a small point worth mentioning in order to correct a misstatement which was made in the previous paper. In a footnote (p. 208) it was there stated that 'in the transformed population the death-rate = 0'. Strictly speaking this would only be approximately true under certain conditions; for, if in the case of the stable canonical population

$$\lambda = 1 + \beta' - \delta',$$

where dashes are attached to the symbols in order to distinguish them from those used above, we have by putting $L_x = 1$ in (2·2) and carrying out the summation,

$$\beta' = \frac{\lambda^{m+1}(\lambda - 1)}{\lambda^{m+1} - 1},$$

and hence
$$\delta' = \frac{\lambda - 1}{\lambda^{m+1} - 1},$$

which will approach zero as λ^{m+1} becomes large. Actually in the numerical example given in the footnote referred to, the value of λ^{m+1} was sufficiently great for δ' to be taken as approximately zero without any very serious error being incurred.

3. THE BIOLOGICAL SIGNIFICANCE OF THE ROW VECTORS

The columns of the matrix M^t are a measure of the contributions made by each age group to the total population at time t. Thus, for example, if there were n_j individuals alive in the age group j to $j+1$ at $t = 0$, the number and age distribution of their living descendants and

survivors at time t could be found by multiplying the elements in the $(j+1)$th column of M^t by n_j, and hence their total contribution to the population at this time is given by n_j times the sum of these elements. It was shown previously (Leslie, 1945, §4) that for values of $t \geqq m-k$, where $x = k$ is the last age group within which reproduction occurs, the last $m-k$ columns of M^t will consist only of zero elements, an expression of the obvious fact that individuals alive in the post-reproductive age groups contribute nothing to the population after they themselves are dead. From the point of view of the contributions made to the future population by the individual age groups, it is the submatrix A^t which is principally of interest. When t becomes very large, A^t can be taken as being proportional to the matrix

$$H^{-1}S_1H = H^{-1}\psi_1\phi_1 H = \xi_1\eta_1,$$

and therefore the sums of the elements in the columns of A^t must be proportional to the row vector η_1. Since a population with an arbitrary age distribution tends ultimately to approach the stable form, provided that the system of age-specific fertility and mortality rates remains constant, it follows that the normalized row vector associated with the dominant latent root provides a measure of the relative contributions per head made to the stable population in the future by the individual age groups. Thus, supposing we have two arbitrary age distributions ξ_x and ξ_y, both subject to the same constant system of age-specific rates, the ratio between the total number of individuals in the two populations would, as time went on, tend to the figure

$$R = \frac{\eta_1\xi_x}{\eta_1\xi_y}.$$

If, instead of regarding ξ_x and ξ_y as two separate populations, we regard them as two components of an age distribution ξ_z, it is thus possible to estimate their relative contributions to the population in the future, subject to the condition that the system of rates represented by the matrix A remains constant.

If, in this expression for R, we put $\xi_y = \xi_1$, the normalized stable vector associated with the dominant root of the matrix, we may write

$$V = \eta_1\xi_x,$$

or, since the angle between two vectors ξ_x and ξ_y, of lengths x and y respectively, is

$$\cos\theta = \frac{\eta_y\xi_x}{yx}, \quad V = x\cos\theta_x,$$

where θ_x is the angle ξ_x makes with the stable vector ξ_1 of unit length. Thus, when ξ_x is the stable form of age distribution $(= c_1\xi_1)$, the quantity V is the same as the length of the vector ξ_x, since $\cos\theta_x = 1$, and when the population is not distributed as to age in the stable form, $0 < V < x$. The rate of increase of V with regard to time is $dV/dt = rV$, since $V(t) = \lambda_1^t V(0)$.

This quantity V appears to be essentially the same as that termed the total reproductive value of a population by Fisher (1930, p. 27). In discussing the equation

$$\int_0^{\infty} e^{-rx}l_x m_x dx = 1,$$

by means of which the inherent rate of increase r is usually calculated, Fisher points out the close analogy between a population increasing geometrically and the growth of capital invested at compound interest. Thus the birth of a child can be regarded as the loaning to him of a life and the birth of his offspring as a subsequent repayment of the debt. Then, 'a unit investment has an expectation of a return $l_x m_x dx$ in the time interval dx, and the present value of this repayment, if r is the rate of interest, is $e^{-rx}l_x m_x dx$; consequently the

Malthusian parameter of population increase is the rate of interest at which the present value of births of offspring to be expected is equal to unity at the date of birth of their parent'. (In this quotation the original symbolism has been changed to that used here; Fisher writes m, the Malthusian parameter, instead of r, and the maternal frequency b_x instead of m_x.) Fisher then goes on to say that 'we may ask, not only about the newly born, but about persons of any chosen age, what is the present value of their future offspring; and if the present value is calculated at that rate determined as before, the question has a definite meaning—To what extent will persons of this age, on the average, contribute to this ancestry of future generations?' He then defines the reproductive value which can be assigned to a person aged x as

$$v_x = \frac{e^{rx}}{l_x} \int_x^\infty e^{-rt} l_t m_t \, dt.$$

Thus, by assigning to each of the n_x persons aged x the appropriate value v_x and summing over all age classes of a given age distribution, a figure which Fisher terms the total reproductive value of the population may be obtained. He also pointed out that this total reproductive value would increase or decrease according to the correct Malthusian rate r.

It was not difficult to show on an actual numerical example that the values of v_x were the same, apart from a scale factor, as the elements of the η_1 row vector after allowing for the fact that the latter refer to a population considered in discrete age groups, whereas the former refer to values of x which vary continuously; and it was evident that the calculation of the quantity V defined above was essentially the same as the calculation of Fisher's total reproductive value of the population.

There is, however, one important point in regard to the argument developed by Fisher which has been quoted. The present value of the repayment $l_x m_x dx$ is taken to be $e^{-rx} l_x m_x dx$, where r is the rate of interest. But, in the case of a population, this estimate of the present value would only be valid if the whole population were increasing at a rate r, and this would only be true when the stable form of age distribution was established. In other words, the reproductive value v_x assigned to a female aged x is the present value of her future daughters only when that female and her daughters are considered as members of a population with a stable age distribution. That this is so may be seen from a numerical example. Let us suppose we are given the age distribution

$$\xi_a = \{81, 21, 5, 1\},$$

which is a stable ξ appropriate to the dominant root $\lambda_1 = 3$ of the numerical matrix A (1·1) defined in the introduction. In one unit of time the population will be $A\xi_a = \lambda_1 \xi_a$ and these individuals will be either survivors or descendants of the original population. Each individual alive in the latter will contribute on the average so many living individuals to the population at $t = 1$, and we wish to assess the present value of that contribution. Consider first of all the solitary female alive in the last age group. In one unit's time this individual will be no longer alive, but she will have contributed $F_3 = 18$ living daughters to the population at that time. The present value of that contribution will therefore be $F_3/\lambda = 6$, and this is the present value which may be attached to each individual alive in this age group of a stable population at any given time. Passing to the five individuals in the next younger age group, $5P_2 = 3$ will be alive in the fourth age group at $t = 1$, and each of these three will be valued then at 6 or a total of 18. They will also have contributed $5F_2 = 90$ daughters. The present total value of the contribution made by these five individuals will be therefore

$$(90+18)/3 = 36.$$

or 7·2 per head. In the same way the 21 individuals in the second age group will each be valued at 3·85714 and the 81 in the first age group at 1 each. These values which have been determined in this way may be written as the row vector

$$\eta_1^* = [1,\ 3{\cdot}85714,\ 7{\cdot}2,\ 6],$$

where an asterisk is attached to the symbol in order to distinguish this vector from the true normalized form for this particular matrix, namely,

$$\eta_1 = \frac{1}{\sqrt{(68)}}[0{\cdot}\dot{3},\ 1{\cdot}28571,\ 2{\cdot}4,\ 2],$$

and it will be noted that $\eta_1^* = 3\,.\sqrt{(68)}\,\eta_1$.

It is clear from this example that this method of assessing the present value of the contribution made by each female aged x to $x+1$ to the population at time $t+1$ is equivalent to determining the present value of her future daughters, and that the valuation can only be carried out in this way when that female and her daughters are considered as members of a stable age distribution. Symbolically the equation which defines the elements y_x $(x = 0, 1, 2, \ldots, k)$ of the vector η^*, and which is equivalent to that given by Fisher for v_x in the continuous case, is

$$y_x = \frac{\sum\limits_x^k \lambda^{-(x+1)} L_x F_x}{L_x \lambda^{-x}},$$

and by an obvious extension to the case of stable 'age distributions' consisting of complex or negative individuals, the stable η_a^* representing the present value of the 'contributions' made by each individual could be calculated similarly for each distinct latent root λ_a of a given matrix A. Moreover, it is evident in each case $y_x = 0$ for all values of $x > k$, the last age group in which reproduction occurs.

The use of these row vectors in the form η_a^* has, however, certain disadvantages, more particularly when it is necessary to compare the total present values of two stable age distributions which are each subject to a different system of rates of death and reproduction. It will be seen from the above equation defining v_x that if the maternal frequency is measured in terms of daughters, we must have in all cases $v_0 = 1$, since

$$\int_0^\alpha e^{-rx} l_x m_x dx = 1 \quad \text{and} \quad l_0 = 1.$$

Similarly in the discrete case, the value of y_0 may be written, making use of the relationship $(P_0 P_1 P_2 \ldots P_x) = L_{x+1}/L_0$,

$$y_0 = \frac{F_0}{\lambda} + \frac{P_0 F_1}{\lambda^2} + \frac{P_0 P_1 F_2}{\lambda^3} + \ldots + \frac{(P_0 P_1 P_2 \ldots P_{k-1}) F_k}{\lambda^{k+1}},$$

which must be equal to unity, since from the characteristic equation of the matrix

$$\lambda^{k+1} - F_0 \lambda^k - P_0 F_1 \lambda^{k-1} - \ldots - (P_0 P_1 \ldots P_{k-2}) F_{k-1} \lambda - (P_0 P_1 \ldots P_{k-1}) F_k = 0.$$

Thus, as exemplified in the numerical illustration given above, the vector η_1^* will always have its first element equal to unity and will in general differ from the normalized η_1 by some scalar factor. The vector η_1^* measures the total value of a stable population on a different scale, or in a different system of units, to those in which the present value is measured by the vector η_1. But the question of the respective units in which a number of such values are expressed might become of importance if two or more stable populations subject to different systems of rates were being compared. Suppose these rates are represented by a number of different

matrices $A_1, A_2, \ldots, A_n$, which will be assumed to be all of the same order. If the series of reproductive values for the individual age groups is taken as the row vector η_1^* appropriate to each of the given matrices, the first element of each vector will necessarily be unity as has been shown above. That the use of these vectors in this form for calculating the total present value may lead to unsatisfactory results for the comparison between two stable populations, can be seen from a simple example. Suppose each element of the numerical matrix A defined in the introduction (1·1), and which we will now call A_1, is divided by a factor of 3. The resulting matrix—say A_2—can then be taken as representing a new system of rates which has a dominant latent root $\lambda_1 = 1$. The stable age distribution $\xi_a = \{81, 21, 5, 1\}$ of A_1 is, however, also a stable ξ_a of A_2 appropriate to this root. If the stable η_1^* for the second matrix is calculated as before the elements will be the same as those given above for the original matrix. The total present value of the population represented by ξ_a would therefore be estimated at the same figure whichever of the two systems of rates it was subject to. If then these were two separate populations with rates A_1 and A_2, which happened to have identical age distributions, a comparison between them by means of the total values calculated in this way is not very informative. The easiest way out of this difficulty would be to use only the normalized η_1 associated with the dominant latent root of each matrix in calculating the total present value of a stable population for the purpose of comparing it with that of another. This procedure allows for any difference in what may be termed the respective scales of the two matrices. For this particular example, the normalized η_1 associated with the root $\lambda_1 = 1$ of the matrix A_2 is

$$\eta_{21} = \frac{0{\cdot}19245}{\sqrt{(68)}}[0{\cdot}\dot{3},\ 1{\cdot}28571,\ 2{\cdot}4,\ 2]$$
$$= 0{\cdot}19245\eta_{11},$$

where the initial of the two suffixes refers to the matrix with which the vector is associated. The total value of a population with an age distribution ξ_a would therefore be 8·2462 if it was subject to the system of rates represented by the matrix A_1, and 1·5870 when subject to A_2. Thus the use of the normalized row vectors instead of the form η_1^* leads to a different value being placed on each of the two populations corresponding to a difference in the systems of rates to which they are respectively exposed.

We may conclude, therefore, that in calculating the total value of a stable population it will in general be preferable to use the normalized stable row vector η_1 and not the form η_1^*. The one form, however, can be readily transformed into the other. For, working in terms of age distributions confined to the prereproductive and reproductive age groups, if the elements of η_1^* are calculated by means of the above equation for y_x, the relationship between η_1^* and η_1 is given by

$$\eta_1 = (P_0 P_1 P_2 \ldots P_{k-1}) \left\{\frac{df(\lambda)}{d\lambda}\right\}^{-\frac{1}{2}} \eta_1^*,$$

where $df(\lambda)/d\lambda$ is the characteristic equation of the matrix differentiated with respect to λ, in which the numerical value of the dominant root is inserted and the square root taken with a positive sign. Thus, for the numerical example which has been used previously in this section, the characteristic equation of the matrix A defined by (1·1) is

$$f(\lambda) = \lambda^4 - 5\lambda^2 - 10\lambda - 6,$$

and

$$\frac{df(\lambda)}{d\lambda} = 4\lambda^3 - 10\lambda - 10.$$

For $\lambda_1 = 3$, we have
$$\left\{\frac{df(\lambda)}{d\lambda}\right\}^{\frac{1}{2}} = \sqrt{(68)},$$
and, since $P_0 P_1 P_2 = \frac{1}{3}$ for this matrix,
$$\eta_1 = \frac{1}{3\sqrt{(68)}}\eta_1^*,$$
corresponding to the difference between these two vectors which was noted above. Although this procedure has been illustrated in terms of the dominant root of the matrix, it can be similarly carried out for any stable η_a^* appropriate to a latent root λ_a. Alternatively, the normalized row vectors may be readily calculated in terms of the canonical population and the matrix $B = HAH^{-1}$ by the methods described in the previous paper (Leslie, 1945, §§ 7 and 8), and transformed back again by means of the relationship $\eta = \phi H$.

If Fisher's total reproductive value of a population is written in terms of vectors as the scalar
$$V = \eta_1 \xi_x = x \cos\theta_x,$$
it follows, as was pointed out earlier in this section, that when the population represented by the vector ξ_x is of the stable form of age distribution, we have $V = x$, the length of ξ_x. The total reproductive value, or the total present value, of a stable population is therefore given by the length of the vector representing the age distribution of the population. Now any population of individuals with a stable form of age distribution ξ_a can be represented as a multiple $c_1\xi_1$ of the normalized stable ξ associated with the dominant root of the matrix, and its associated vector η_a as a multiple $\bar{c}_1\eta_1$ of the normalized η_1, the square of the length of ξ_a being given by $\eta_a\xi_a$. We may thus regard the vector $\eta_a = \bar{c}_1\eta_1$, which is associated with the vector $\xi_a = c_1\xi_1$, as the representative of the population in terms of the individual present values according to age, just as the vector ξ_a is the representative of the population in terms of numbers according to age. Although we have been here considering only the total present value of a population of real positive individuals distributed as to age in the stable form, which must necessarily involve only one of the stable η or ξ for a given matrix, there is little difficulty from the mathematical point of view in considering 'populations' consisting of negative or complex individuals, and we may extend the arguments used for the real case so as to include all the stable vectors for the matrix. Thus, the length of any stable vector, ξ_a say, which fulfils the condition $A\xi_a = \lambda_a\xi_a$, can be regarded as the total present value of the 'population' represented in terms of numbers by ξ_a and in terms of individual present values by its associated vector η_a.

Since any arbitrary age distribution of real individuals ξ_x can be regarded as the sum of one or more mutually orthogonal stable ξ, viz.
$$\xi_x = c_1\xi_1 + c_2\xi_2 + \ldots + c_{k+1}\xi_{k+1},$$
and its associated vector η_x similarly as the sum of a number of associated stable η
$$\eta_x = \bar{c}_1\eta_1 + \bar{c}_2\eta_2 + \bar{c}_3\eta_3 + \ldots + \bar{c}_{k+1}\eta_{k+1},$$
and since the total present value of each of the component stable vectors is given by the length of that vector, namely $\sqrt{(\bar{c}_a c_a)}$, the total present value of the resultant ξ_x will be given by $\sqrt{(\sum \bar{c}_a c_a)}$, which is the length of the vector ξ_x.

The row vectors which were originally introduced into this theoretical discussion solely for mathematical reasons are thus not entirely without interest from the biological point of view. The uniquely determined vector η_x which was assumed to be associated with each ξ_x is a measure of the present value of the contribution made to future generations by an

individual aged x to $x+1$ when that individual is considered as a member of a population with an age distribution ξ_x. The row vectors appear to form a more generalized system of weights or values which we attach to an individual aged x to $x+1$ than the reproductive values v_x defined by Fisher. The latter are represented by a single member of this class of vectors, though one of particular importance owing to its association with the dominant root of the matrix.

Finally there is one further row vector which is very easily calculated for a given system of age-specific fertility and mortality rates, and which on occasion may be useful in studying the comparative fertility of different populations. The net reproduction rate,

$$R_0 = \int_0^\infty l_x m_x dx,$$

in addition to its usual meaning, may also be defined as the expected number of daughters which will be born on the average by a female now aged 0 during the remainder of her lifetime. It is in fact a figure which is analogous to the expectation of life at birth, only in terms of future daughters. Now, in addition to the newly born, we may also enquire what this expected number of daughters will be in the case of a female alive at any age x. Clearly this figure is given by,

$$u_x = \frac{1}{l_x}\int_x^\infty l_x m_x dx,$$

with $u_0 = R_0$. Similarly, in the discrete case, we may consider an η row vector of which the elements z_x $(x = 0, 1, 2, \ldots, k)$ are

$$z_x = \frac{1}{L_x}\sum_x^k L_x F_x,$$

and it will be found that this is merely a multiple of the η_1^* vector appropriate to the dominant root $\lambda_1 = 1$ of the matrix for a stationary population which is obtained by dividing each of the F_x figures in the first row of the matrix A by the net reproduction rate.

4. THE TOTAL REPRODUCTIVE VALUE OF A POPULATION AND THE LENGTH OF A VECTOR

It appears from the foregoing discussion that the elements of the normalized row vector η_1 can be regarded from two slightly different points of view. On the one hand they provide a measure of the relative contributions per head made by each age group to the stable population in the future, and this property arises from the fact that the sums of the columns of the matrix A^t can be taken as proportional to the elements of this vector when t becomes very large. On the other hand this vector is also associated with the column vector ξ_1 representing the stable age distribution appropriate to a given matrix, and in this sense its elements are a measure of the present value of the contribution made to future generations by an individual aged x to $x+1$ when that individual is considered as a member of a population with a stable age distribution. This difference is of importance in making any practical use of Fisher's total reproductive value of a population, which is defined here as $V = \eta_1 \xi_x$, where ξ_x is an arbitrary age distribution.

Thus, if we have two populations ξ_x and ξ_y both of which are subject to the same system of rates A, or alternatively if ξ_x and ξ_y are two subdivisions of one population subject to A, we can calculate for each the total reproductive values V_x and V_y, and determine the ratio $R = V_x/V_y$. This quantity, as was shown at the beginning of the previous section, is the ratio at time t, when t becomes very great, of the total number of individuals in the two populations which at $t = 0$ had the age distributions ξ_x and ξ_y. But the quantity R cannot be interpreted in this way when the two populations are not subject to the same system of rates.

Again, if a population happens to have a stable form of age distribution ξ_a, then $V = \eta_1 \xi_a = a$, the length of the vector ξ_a and this figure represents the total present value of the stable population ξ_a. But, apart from the case when an arbitrary ξ_x is of the stable form, it is difficult to define the meaning of V simply by itself in any precise biological terms. From the mathematical point of view, when an arbitrary ξ_x is expanded in terms of the stable ξ, and

$$\xi_x = c_1 \xi_1 + c_2 \xi_2 + \dots + c_{k+1} \xi_{k+1},$$

we have

$$\eta_x \xi_x = \sum_{a=1}^{k+1} \bar{c}_a c_a,$$

which is the same thing as x^2, the square of the length of the vector ξ_x. Then it can be seen that since $V = \eta_1 \xi_x = c_1 = x \cos \theta_x$, the calculation of Fisher's total reproductive value is essentially the determination of one component of a set of mutually orthogonal sums of squares which together make up the total sum of squares represented by x^2. Thus $V^2 = c_1^2$ which is the first term in $\eta_x \xi_x = \sum_a \bar{c}_a c_a$, since c_1 is necessarily a real positive number.

The two methods of valuation which have been mentioned here are the calculation of the length of the vector ξ_x representing the age distribution of the population, and the calculation of the total reproductive value V. Which of these two figures is the more important from the point of view of assessing the state of a population subject to a given system of fertility and mortality rates is a matter for discussion and further investigation. Certainly the total reproductive value V is a figure which is the more easily determined. It requires only a knowledge of the row vector η_1 associated with the dominant root of the matrix representing the given system of rates to which the population is subject. On the other hand the calculation of the length of the vector ξ_x, is much more complicated. For, in order to arrive at the associated vector $\eta_x = \xi_x' HGH$, it is necessary to know the numerical values of the elements of the matrix G, and hence HGH, which in turn cannot be computed unless all the latent roots of the matrix A are known. Thus, purely from the practical point of view, the calculation of the total reproductive value $V = \eta_1 \xi_x$ offers a number of advantages and, within the limitations set out above, this figure may prove useful in comparing one population with another.

It is perhaps worth mentioning in passing one further type of problem. If the length of the vector ξ_x is regarded as the present value of the population when it is subject to a particular system of fertility and mortality rates, it may be of interest on occasion to consider the maximum or minimum of the quadratic form $\xi' HGH \xi$ given one or more restrictive conditions. Thus, for example, we might consider the problem of determining the column vector ξ_s which would give rise to the minimum total value when the sum of its elements was equal to a number N. If n_x $(x = 0, 1, 2, \dots, k)$ are the elements of ξ_s and the symbol $\{1\}$ represents a column vector of $(k+1)$ units, we have, after differentiating with respect to the n_x and introducing a Lagrange multiplier λ,

$$HGH\xi_s - \lambda\{1\} = 0,$$
$$\Sigma n_x = N,$$

a set of $(k+2)$ equations for determining the values of n_x which will make the length of the vector ξ_s a minimum subject to the restrictive condition imposed. It will be seen from these equations that the solution of this problem is equivalent to that of determining the column vector ξ_s which will have all the elements of its associated row vector η_s the same value. Thus, by reversing the process, and starting with an arbitrary row vector of $(k+1)$ units, it

follows that the required column vector is proportional to the sums of the columns of the matrix $H^{-1}G^{-1}H^{-1}$. As an example of the type of vector which has the minimum value, the solution of these equations in the case of the simple 4×4 matrix given in the introduction was for $N = 108$,
$$\xi_s = \{84{\cdot}5686,\ 17{\cdot}4054,\ 4{\cdot}5059,\ 1{\cdot}5201\},$$
whereas the stable population of 108 individuals was
$$\xi_1 = \{81,\ 21,\ 5,\ 1\}.$$
This problem has been considered here in terms of the vector of shortest length, without imposing the full restrictive conditions which strictly speaking would be necessary when considering a population of living individuals, namely that the elements n_x of the column vector are positive integers with $\Sigma n_x = N$. But the vector ξ_s in this example consists of positive elements and may be taken as representing, in the case of this numerical system, the type of proportionate age distribution which would give rise to the minimum value. Actually the difference between the two distributions ξ_s and ξ_1 is not very marked in this example. The square of the length of the stable vector is 68, while that of the vector of shortest length is 64·4. But that this difference between the total values does correspond to a difference between the properties of the two age distributions may be seen by operating on each of them with the matrix A and determining the total number of individuals in the two populations at successive intervals of time. The numbers in the population which starts with an age distribution ξ_s will always be lower than those in the population starting with the stable form ξ_1, until ultimately there would be about 5·3 % fewer individuals in the former than in the latter.

5. THE LIMITED TYPE OF POPULATION GROWTH

Hitherto it has been assumed that the system of age-specific fertility and mortality rates represented by the matrix A remains constant, and that therefore the population increases geometrically to an unlimited extent at a rate $dN/dt = rN$, when the stable age distribution is established. The next case which is usually considered in population mathematics is that of the logistic population, where the rate of increase in numbers is defined by the differential equation
$$\frac{dN}{dt} = (r - aN)N,$$
r and a being constants > 0, from which the well-known result follows that such a population will approach asymptotically an upper limit to the numbers given by $K = r/a$, according to the equation
$$N = \frac{K}{1 + Ce^{-rt}}.$$
It is therefore of interest to consider in terms of matrices and vectors the type of population growth in which the system of rates is dependent on the number of individuals present in the population at a given time.

Suppose that the system of rates to which a population is exposed when no limitations are placed upon the growth in numbers is represented by the matrix A with a dominant latent root λ_1. This might be called the optimum system of rates for the particular species or genetic stock. When the population is increasing in a limited environment let us suppose that at time t there is an age distribution $\xi(t)$ consisting of a total number $N(t)$ of individuals, and that at this time the elements of A are altered so that we have a new matrix A_t with a dominant latent root $\lambda_1/q(t)$, where $q(t)$ is dependent on $N(t)$. Then the age distribution of

the population at time $t+1$ will be given by $A_t \xi(t) = \xi(t+1)$, and the process can be obviously extended so that at time $t+1$ we have a matrix A_{t+1} with a dominant root $\lambda_1/q(t+1)$, $q(t+1)$ depending on $N(t+1)$, and so on. At each integral value of t, therefore, the original inherent rate of increase $r = \log_e \lambda_1$ will in general change to a new rate $r' = \log_e (\lambda_1/q)$, where q is some function of N, the number of individuals present in the population.

The changes which are thus assumed to occur in the optimum age-specific rates of fertility and mortality represented by the matrix A might take place in an innumerable variety of different ways. But, from the theoretical point of view, there are two extreme cases which are particularly of interest; on the one hand, when the decrease in the optimum rate of increase is due to a lowered degree of fertility, while the age-specific death-rates remain the same: and on the other when it is due to an increased rate of mortality and fertility remains constant. Even under these simplified conditions it is necessary to make some assumption as to the way in which the rates are actually affected, and in order to define the problem in concrete terms, it will be assumed here that the changes which occur either in the degree of fertility or in that of mortality are due to the operation of a factor which is independent of age. In addition one further type of change in the rates of fertility, involving a factor which increases geometrically with age, will be mentioned in passing. For simplicity the two main cases will be considered separately.

(*a*) *Mortality affected by a factor independent of age, fertility remaining constant*

If l_x and m_x are respectively the life table and fertility table for a population living under optimum conditions where no limitations are placed upon the growth in numbers, the inherent rate of increase (r) of the population is defined by

$$\int_0^\infty e^{-rx} l_x m_x \, dx = 1,$$

and the stable age distribution (c_x) and the stable birth-rate (b) by

$$c_x = b e^{-rx} l_x, \quad \frac{l}{b} = \int_0^\infty e^{-rx} l_x \, dx.$$

If now a force of mortality (γ) which is independent of age is superimposed on the original force of mortality (μ_x), represented by the optimum life table l_x, the new life table l'_x will be given by

$$\frac{1}{l'_x}\frac{dl'_x}{dx} = -(\gamma + \mu_x) \quad \text{or} \quad l'_x = e^{-\gamma x} l_x;$$

and, if the original fertility table remains unaltered, the new inherent rate of increase will be $r' = r - \gamma$. The stable age distribution (c'_x) and stable birth-rate (b') of the population when it is subject to this new life table will then be

$$c'_x = b' e^{-r'x} l'_x, \quad \frac{l}{b'} = \int^\infty e^{-r'x} l'_x \, dx;$$

and it follows, since $l'_x = e^{-\gamma x} l_x$ and $r' = r - \gamma$, that $1/b' = 1/b$ and $c'_x = c_x$. The imposition of a force of mortality independent of age on a given life table thus leaves the original stable age distribution and stable birth-rate unchanged.

Similarly in terms of matrices, if A is the matrix representing the age-specific rates of fertility and mortality for a population living under optimum conditions, we are led to consider the matrix $q^{-1}A$ in which each element of the original matrix A is divided by a scalar q. Approximately, in the discrete case, this is equivalent to imposing on the original

life table a force of mortality which is independent of age. Then in the reduction of $q^{-1}A$ to rational canonical form, if

$$H_q = \begin{bmatrix} (P_0P_1P_2\dots P_{k-1})q^{-k} & \cdot & \dots & \cdot & \cdot & \cdot \\ \cdot & (P_1P_2\dots P_{k-1})q^{-(k-1)} & \dots & \cdot & \cdot & \cdot \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \cdot & \cdot & \dots & (P_{k-2}P_{k-1})q^{-2} & \cdot & \cdot \\ \cdot & \cdot & \dots & \cdot & (P_{k-1})q^{-1} & \cdot \\ \cdot & \cdot & \dots & \cdot & \cdot & 1 \end{bmatrix},$$

the first row of $B_q = H_q(q^{-1}A)H_q^{-1}$ is

$$F_0q^{-1},\quad P_0F_1q^{-2},\quad P_0P_1F_2q^{-3},\quad \dots,\quad (P_0P_1P_2\dots P_{k-1})F_kq^{-(k+1)},$$

the remaining elements consisting in the usual way of a series of units in the principal subdiagonal.

The characteristic equation of the matrix B_q is

$$\lambda^{k+1} - F_0q^{-1}\lambda^k - P_0F_1q^{-2}\lambda^{k-1} - \dots - (P_0P_1\dots P_{k-2})F_{k-1}q^{-k}\lambda - (P_0P_1\dots P_{k-1})F_kq^{-(k+1)} = 0$$

while that of the original matrix $B = HAH^{-1}$ is obtained by putting $q = 1$. Comparing these two equations term by term it will be seen that the latent roots of B_q are merely those of B each divided by the factor q. Thus, in terms of the canonical population, the stable age distribution appropriate to the dominant latent root λ_1/q of the matrix B_q may be taken as a multiple of the vector

$$\psi_1 = \left\{\left(\frac{\lambda_1}{q}\right)^k,\ \left(\frac{\lambda_1}{q}\right)^{k-1},\ \dots,\ \left(\frac{\lambda_1}{q}\right),\ 1\right\}:$$

and since $\xi_1 = H_q^{-1}\psi_1$,

$$\xi_1 = \{(P_0P_1P_2\dots P_{k-1})^{-1}\lambda_1^k,\ (P_1P_2\dots P_{k-1})^{-1}\lambda_1^{k-1},\ \dots,\ (P_{k-1})^{-1}\lambda_1,\ 1\},$$

which is the same as the $\xi_1 = H^{-1}\psi_1$ appropriate to the root λ_1 of the original matrix A. Moreover, since the time which it takes for an arbitrary ξ_x to approach the stable form of age distribution associated with the dominant root of the matrix will depend on the ratios of this root to the other roots of the matrix, as may be seen from the expansion of $\xi_x(t)$ at time t in terms of the stable ξ,

$$\xi_x(t) = c_1\lambda_1^t\xi_1 + c_2\lambda_2^t\xi_2 + \dots + c_{k+1}\lambda_{k+1}^t\xi_{k+1},$$

it follows that a population with any arbitrary form of age distribution which is subject to the matrix $q^{-1}A$ will approach the stable form at the same rate for all values of q. This result is of interest in the theoretical study of wild mammalian populations, since it might be assumed, at least as a first approximation, that any increase of mortality due to predation, hunger, etc., falling on some optimum system of age-specific death-rates could be represented by a factor which tended to be independent of age.

If then we consider at time t a population with an age distribution $\xi(t)$ which is subject to the system of rates represented by the matrix $q^{-1}A$, and we regard q as some function of N, the number of individuals present in the population at time t, we might put as a first approximation

$$q = \alpha + \beta N.$$

For the stationary state we must have $q = \lambda_1$, the dominant latent root of the matrix A,

and in addition, as N tends to zero, q must approach 1. When the dominant latent root of $q^{-1}A$ is equal to unity, the condition for a stationary population,

$$N = \frac{\lambda_1 - 1}{\beta} = K,$$

and therefore we may write
$$q = 1 + \frac{(\lambda_1 - 1)\,N}{K}.$$

Then, assuming at time t there are $N(t)$ individuals distributed as to age in the stable form of distribution (ξ_a) for the matrix $q^{-1}A$, which distribution is the same for all values of q as has been shown above,

$$q^{-1}A\xi_a(t) = \xi_a(t+1) = \frac{\lambda_1 \xi_a(t)}{q},$$

or

$$N(t+1) = \frac{\lambda_1 N(t)}{1 + \dfrac{\lambda_1 - 1}{K} N(t)},$$

and

$$\frac{K - N(t+1)}{N(t+1)} = \lambda_1^{-1}\left\{\frac{K - N(t)}{N(t)}\right\},$$

which, as $\log_e \lambda_1 = r$, is the same thing as the logistic type of population growth,

$$N = \frac{K}{1 + Ce^{-rt}}.$$

Thus, when fertility remains constant and mortality is affected by a factor which is independent of age, this factor being regarded as a simple linear function of the numbers present in the population at time t, the total number of individuals in the population will increase according to the logistic form of population growth, provided that the age distribution of the population at $t = 0$ is the stable form appropriate to the dominant latent root of the matrix A. But, when this condition is not fulfilled, and the initial age distribution is not of the stable form, there may be quite considerable departures from the curve given by this actual logistic equation. The form of the curves representing the total number of individuals at successive intervals of time will, however, still tend to be S-shaped, and in some cases there is little doubt that a logistic type of equation could be fitted empirically to the data over a considerable portion of the total curve. The type of variation which might be expected in these growth curves owing to a departure from the stable form of age distribution is illustrated in the following simple examples.

Suppose that an entirely imaginary population, which can be considered in four age groups, is subject to the optimum system of rates of death and reproduction represented by the matrix defined originally in the introduction,

$$A = \begin{bmatrix} 0 & 6{\cdot}4286 & 18 & 18 \\ 0{\cdot}7778 & 0 & 0 & 0 \\ 0 & 0{\cdot}7143 & 0 & 0 \\ 0 & 0 & 0{\cdot}6000 & 0 \end{bmatrix},$$

which has a dominant root $\lambda_1 = 3$, or $r = 1{\cdot}09861$, and suppose that for the matrix $q^{-1}A$,

$$q = 1 + 0{\cdot}000185185N.$$

where N is the number of individuals in the population at integral values of time t. When

$q = 3$, the stationary state, $N = 10800$; and at $t = 0$ let there be 108 individuals present in the population. These conditions are fulfilled by the logistic equation

$$N = \frac{10800}{1 + 99e^{-1\cdot 09861t}}.$$

If at $t = 0$ we consider three different age distributions each consisting of 108 individuals and represented by the vectors

$$\xi_a = \{81, 21, 5, 1\}, \quad \xi_s = \{85, 17, 4, 2\}, \quad \xi_x = \{0, 0, 108, 0\},$$

where ξ_a is a stable age distribution of the matrix A, ξ_s the vector of shortest length given in the previous section and expressed to the nearest integer, and ξ_x a very skew form of age distribution in which all the individuals are concentrated in an age class for which fertility is high, the age distributions and therefore the total number in each population can be readily calculated by successive applications of the matrix $q^{-1}A$. The following are the results obtained in each case, together with the values of N calculated from the logistic equation

Values of N

t	From logistic	Initial age distribution		
		ξ_a	ξ_s	ξ_x
0	108·0	108	108	108
1	317·6	318	292	1970
2	900·0	901	844	1930
3	2314·3	2316	2215	6199
4	4860·1	4862	4660	8423
5	7673·7	7675	7540	9389
6	9508·7	9509	9433	10694
7	10332·3	10332	10298	10609
8	10639·5	10641	10628	10741
9	10745·9	10745	10742	10804
10	10781·9	10781	10780	10781

which are given in the first column. It will be seen that in the case of the stable age distribution ξ_a the values of N follow those calculated from the logistic equation, apart from small discrepancies at times in the last figure due to errors of rounding off. (The elements of the vector $\xi(t+1) = q^{-1}A\xi(t)$ were in each case expressed to the nearest whole number.) In the case of ξ_s, an age distribution which does not differ very greatly from the stable form, the numbers lie below those for the initial distribution ξ_a until $t = 10$, the stable age distribution being approximately established in this population round about $t = 7$; while for ξ_x the numbers are very erratic owing to the very skew form of the initial distribution leading to a very rapid increase in numbers during the early stages. The stable form of age distribution was approximately established in this last population at $t = 10$. It is evident from these examples that the initial form of age distribution may have a marked effect on the course of development followed by a population which inherently is increasing towards some upper limit according to the type of growth in numbers assumed here.

The initial number of individuals in these three examples is small relative to the upper limit of $K = 10800$, so that even a thoroughly skew form of distribution such as ξ_x has time

in which to approach the stable form of age distribution before the upper limit in numbers is achieved. Actually in the case of ξ_x the stable form is not established before $t = 10$, and a tendency to overshoot the upper limit will be noticed before that time. If the initial number of individuals had been chosen much greater relative to K this tendency would only have been emphasized. An extreme case would have been to assume that the initial number of individuals in each of the three examples was equal to 10800. Then it is evident that whereas the population represented by ξ_a, the stable form, would have remained constant at the same figure, those represented by ξ_s and ξ_x would vary on either side of the upper limit to begin with and would tend to approach the steady state by a series of damped oscillations as the stable age distribution was in the process of being established.

(b) Fertility affected by a factor independent of age, mortality remaining constant

This problem raises a number of difficulties not all of which have been satisfactorily resolved. But, before considering the main problem as defined here, namely when fertility is affected by a factor independent of age, there is another case which arises from the foregoing discussion, and which is perhaps worth mentioning. The canonical matrix

$$B_q = H_q(q^{-1}A)H_q^{-1}$$

defined above, is when written in full, to take a simple example of a 4×4 matrix,

$$B_q = \begin{bmatrix} F_0q^{-1} & P_0F_1q^{-2} & P_0P_1F_2q^{-3} & P_0P_1P_2F_3q^{-4} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix};$$

and it can be seen that in addition to being the canonical form of $q^{-1}A$, B_q is also the canonical form of

$$A_q = \begin{bmatrix} F_0q^{-1} & F_1q^{-2} & F_2q^{-3} & F_3q^{-4} \\ P_0 & 0 & 0 & 0 \\ 0 & P_1 & 0 & 0 \\ 0 & 0 & P_2 & 0 \end{bmatrix},$$

the diagonal matrix H of the transformation HA_qH^{-1} having elements $h_{11} = P_0P_1P_2$, $h_{22} = P_1P_2$, $h_{33} = P_2$, $h_{44} = 1$. This matrix A_q can be regarded as representing some system of age-specific rates in which an original level of fertility included in the F_x figures has been affected by a factor which increases geometrically with age, and as before this factor q might be taken as being linearly related to N, the number of individuals present in the population at time t. But in contradistinction to the matrix $q^{-1}A$, the age distribution of a population subject to the matrix A_q will no longer remain stable. For suppose that in terms of the canonical population the stable age distribution associated with the dominant root λ_1/q of A_q is

$$\psi_1 = \left\{\left(\frac{\lambda_1}{q}\right)^k, \left(\frac{\lambda_1}{q}\right)^{k-1}, \ldots, \left(\frac{\lambda_1}{q}\right), 1\right\},$$

the transformation $\xi_1 = H^{-1}\psi_1$ gives

$$\xi_1 = \left\{(P_0P_1P_2\ldots P_{k-1})^{-1}\left(\frac{\lambda_1}{q}\right)^k, (P_1P_2\ldots P_{k-1})^{-1}\left(\frac{\lambda_1}{q}\right)^{k-1}, \ldots, P_{k-1}^{-1}\left(\frac{\lambda_1}{q}\right), 1\right\},$$

which is not the same as the ξ_1 associated with the dominant root λ_1 of the original matrix A.

If then at time t a population happened to have the stable form of age distribution appropriate to the matrix $A_q(t)$, it will not in general have the stable form of distribution at $t+1$ appropriate to $A_q(t+1)$, except in the case of the stationary population with $N = K$ and $\lambda_1/q = 1$, when the life table age distribution is established.

By extending this argument for the matrix A_q to the perfectly general case, it can be seen that the age structure of a population will be constantly changing when the degree of fertility is affected and the life table remains constant, until in the terminal stages of its growth the population approaches the stationary state. This is, of course, essentially the same type of changing age distribution as that shown to occur by Lotka (1931) in the case of a population growing in numbers according to the logistic law with a constant form of life table. A numerical example is given later of a population subject to the matrix A_q when q is taken as a simple linear function of N.

Although, biologically speaking, it is not impossible for fertility to be affected by a factor which increases geometrically with age and which depends on the number of individuals present in the population at a given time, it is perhaps of greater interest to consider the case in which the fractional decrease in fertility is the same at all ages. In other words, it is necessary to consider the matrix A_s, say, in which the elements in the first row of a matrix A representing the optimum rates of death and reproduction are each divided by a factor s, so that

$$A_s = \begin{bmatrix} F_0 s^{-1} & F_1 s^{-1} & F_2 s^{-1} & \dots & F_k s^{-1} \\ P_0 & . & . & \dots & . \\ . & P_1 & . & \dots & . \\ \dots & \dots & \dots & \dots & \dots \\ . & . & . & P_{k-1} & . \end{bmatrix}.$$

Now, if the $(x+1)$th element in the first row of the canonical form $B = HAH^{-1}$ is written as

$$(P_0 P_1 P_2 \dots P_{x-1} F_x) = f_x,$$

the characteristic equation of the original matrix A is

$$\lambda^{k+1} - \left(\sum_{x=0}^{k} f_x \lambda^{k-x}\right) = 0,$$

and that of the matrix A_s is $$\lambda^{k+1} - \left(s^{-1} \sum_{x=0}^{k} f_x \lambda^{k-x}\right) = 0.$$

If the real positive root of the first equation is λ_1, the real positive root of the second can be written as λ_1/q, and the inherent rate of increase of a population subject to the system of rates A_s will be $r' = \log_e(\lambda_1/q)$. Since we are considering as before the case when q is a function of N, say

$$q = 1 + \frac{\lambda_1 - 1}{K} N,$$

it is necessary, in order to solve the problem of a population in which fertility is affected by a factor independent of age, that s should be expressed as a function of q.

This point proved to be rather troublesome, and the following solution needs a much fuller investigation than it has received here. It depends on the relation between the first row of the canonical form $B = HAH^{-1}$ and the $L_x m_x$ column which was touched on in the previous paper (Leslie, 1945, § 6). It is evident that the division of the elements in the first row of the

matrix A or B by a scalar s is the same thing as dividing the maternal frequency figures (m_x) by the same quantity. The original net reproduction rate, $R_0 = \int_0^\infty l_x m_x dx$, will therefore become R_0/s. Now, in the solution of the equation

$$\int_0^\infty e^{-rx} l_x m_x dx = 1,$$

we have
$$\log_e R_0 = m_1 r - \frac{m_2}{2!} r^2 + \frac{m_3}{3!} r^3 - \frac{m_4 - 3m_2^2}{4!} r^4 + \dots, \tag{5·1}$$

where
$$m_1 = \int_0^\infty x l_x m_x dx \bigg/ \int_0^\infty l_x m_x dx,$$

and m_n $(n = 2, 3, 4, \dots, n)$ is the nth moment about this mean. When the maternal frequency is divided by s the moments of the distribution will not be affected, but the value of r will change to a new value r', and

$$\log_e (R_0/s) = m_1 r' - \frac{m_2}{2!} r'^2 + \frac{m_3}{3!} r'^3 - \frac{m_4 - 3m_2^2}{4!} r'^4 + \dots. \tag{5·2}$$

The moments are usually calculated by treating the $L_x m_x$ figures $\left(L_x = \int_x^{x+1} l_x dx\right)$ as a frequency distribution, the individual frequencies being regarded as centered at the mid-point of each age group. Alternatively they are sometimes calculated from $l_x m_x$, where l_x is the value of the usual life table function taken at each midpoint. When a system of rates is expressed in the form of a matrix the elements of the first row of the canonical form $B = HAH^{-1}$ are not the same as the $L_x m_x$ figures. But it was found (Leslie, 1945, § 6) that the sum of these elements was equal to the net reproduction rate and that if each element $(P_0 P_1 P_2 \dots P_{x-1} F_x)$ was regarded as centered at the age of $x+1$, the mean and semiinvariants of the distribution were the same as those obtained from the $L_x m_x$ column.

These relationships suggested a possible way of relating s to q. If for the matrix A, with a dominant latent root $\lambda_1 = e^r$, the sum of the elements in the first row of $B = HAH^{-1}$ is equal to R_0, and if the dominant latent root of the matrix A_s is $\lambda_1/q = e^{r'}$, we might, as a first approximation, take only the first terms in each of the equations (5·1) and (5·2), and put

$$\log_e R_0 = m_1 r, \quad \log_e (R_0/s) = m_1 r',$$

and
$$\frac{r'}{r} \log_e R_0 = \log_e (R_0/s),$$

or, since
$$\frac{r'}{r} = 1 - \frac{\log_e q}{r},$$

$$\log_e s = \frac{\log_e R_0}{r} \log_e q. \tag{5·3}$$

For a greater degree of accuracy the first two terms could be taken as (5·1) and (5·2), viz.

$$\log_e R_0 = m_1 r - \frac{m_2}{2} r^2, \quad \log_e (R_0/s) = m_1 r' - \frac{m_2}{2} r'^2;$$

and
$$\frac{r'}{r}\left(1 + \frac{r - r'}{\dfrac{2m_1}{m_2} - r}\right) \log_e R_0 = \log_e (R_0/s).$$

From which, putting $\log_e q = w$ and $\frac{2m_1}{m_2} - r = c$,

$$\log_e s = -\frac{\log_e R_0}{rc}\{(r-c)w - w^2\}. \tag{5·4}$$

For a greater degree of accuracy still, further terms on the right-hand side of (5·1) and (5·2) could be included, though the algebra tends to become somewhat tedious. Presumably the number of terms which it would be necessary to include in any particular case would depend on the magnitude of r and upon the form of the distribution relating net fertility to age. Actually in the elementary numerical example which has been used here so far, equation (5·4) appears to be fairly accurate. Thus the characteristic equation of A with $\lambda_1 = 3$ and $R_0 = 21$ is

$$\lambda^4 - 5\lambda^2 - 10\lambda - 6 = 0.$$

Dividing these numerical coefficients by $s = 5$, for example,

$$\lambda^4 - \lambda^2 - 2\lambda - 1{\cdot}2 = 0,$$

of which the real positive root is $\lambda_1/q = 1{\cdot}63476$, or $q = 1{\cdot}83513$. The values of m_1 and m_2 were 3·04762 and 0·52154 respectively, and equation (5·4) was in common logarithms

$$\log s = 2{\cdot}48366 \log q + 0{\cdot}60263 (\log q)^2. \tag{5·5}$$

For $q = 1{\cdot}835$, the estimated value of s is 4·975, whereas the true value is $s = 5$. If s is estimated from equation (5·3) for $q = 1{\cdot}835$ the value is 5·379, so that the second degree equation in $\log q$ is an improvement on the first and gives a reasonably close approximation to s for values lying in this region. It will be noted that if $q = 3$, $s = 21$ from this second degree equation (5·5), as it should do.

In order to compare the operational effect of the matrix A_s with that already determined for $q^{-1}A$, two examples are given below for the initial age distributions

$$\xi_a = \{81, 21, 5, 1\}, \quad \xi_x = \{0, 0, 108, 0\},$$

ξ_a being the stable age distribution of 108 individuals for the matrix A, and ξ_x the same form of skew distribution used previously. As before, q was taken as

$$q = 1 + 0{\cdot}000185185N,$$

and the appropriate value of s at each stage was calculated by means of equation (5·5). In addition one example is given of the operation of the matrix A_q in which fertility is affected by a factor which increases geometrically with age, taking ξ_a as the initial distribution. The results were as follows:

Values of N

t	From logistic	Matrix A_q ξ_a	Matrix A_s ξ_a	Matrix A_s ξ_x
0	108·0	108	108	108
1	317·6	312	312	1915
2	900·0	867	867	1976
3	2314·3	2118	2115	5603
4	4860·1	4194	4120	7315
5	7673·7	6659	6393	8616
6	9508·7	8857	8362	10464
7	10332·3	10268	9696	10369
8	10639·5	10876	10384	10695
9	10745·9	10984	10673	10901
10	10781·9	10900	10766	10715

Comparing the two cases in which the initial distribution was of the stable form ξ_a with the figures derived from the logistic curve, it will be seen that in both cases the numbers of individuals are less than those for the logistic particularly in the early stages of development. Broadly speaking, however, all these three curves are similar in their general outlines, though there is an obvious tendency in the case of the matrix A_q for the population to overshoot the upper limit of $N = 10800$ in the later stages. Similarly, in the case of the initial distribution ξ_x and the matrix A_s, the course of events is not very different from that for the previous example with this distribution, when it was assumed that mortality was changing and fertility remained constant, though, again here, the numbers of individuals are less when fertility is changing and mortality remains the same. The chief difference between these examples and those given previously lies, of course, in the forms of the age distribution. When the matrix $q^{-1}A$ was assumed to be in operation, the ultimate age distribution to which all populations would tend, whatever their initial conditions and numbers might be, was

$$\xi = \{8100,\ 2100,\ 500,\ 100\};$$

whereas, both for the matrix A_q and A_s, the stationary age distribution of 10800 individuals is

$$\xi = \{4050,\ 3150,\ 2250,\ 1350\};$$

and throughout the whole course of development of each population an approach is being made to one or other of these very different distributions.

Although the two extreme cases of either fertility or mortality changing through the operation of a factor which is independent of age have been considered here separately, there should be little difficulty in extending the methods so as to include the case where both fertility and mortality are affected in varying degrees at the same time. Thus, we might consider the scalar q of the dominant latent root λ_1/q at time t as being the product, $q = uv$, of two factors, one of which, u say, represents an increase in mortality independent of age, and the other v represents the effect of a decrease in fertility at all ages by means of the factor s. Various possibilities then arise, depending on whether the ratio u/v was regarded as a constant, or as varying in some predetermined manner. However, these questions have not been gone into any further at present.

It will be noticed that the problem considered in this section of a growing population subject to a changing degree of fertility and a constant life table is not precisely the same as that discussed by Lotka (1931). In the first part of that paper Lotka showed how the birth-rate, death-rate, age distribution and inherent rate of increase of such a population would change when the total number of individuals in the population increased according to the logistic law. Here no assumption is made as to the way in which the number of individuals is increasing, but it is assumed that at equal intervals of time, which intervals in practice can be made as small as we please according to the degree of accuracy required, the inherent rate of increase of the population $r' = \log_e (\lambda_1/q)$ is dependent on the number of individuals (N) present at time t, and, as a first approximation, q has been taken as a linear function of N. The most important feature of this form of population growth is the marked effect which the initial age distribution and numbers have on the subsequent course of development of the population. Only in one case, namely when mortality is increased owing to the operation of a factor independent of age, fertility remaining constant, and when the initial age distribution is of the stable form appropriate to the matrix A, is the true logistic form of growth in numbers realized. However, the result of operating on a not too abnormal initial distribution with either of the matrices $q^{-1}A$, A_q or A_s is, broadly speaking, a very similar type

of S-shaped curve, if the initial numbers are small relative to the upper limit K, and in some cases there is little doubt that a logistic equation could be fitted empirically to such a series of points, more particularly when the figures for the total number of individuals are not available over the complete range of development of the population. But, in general, we shall have for a given matrix A and a given value of K in the equation $q = 1+(\lambda_1 - 1)N/K$, a family of S-shaped or partially S-shaped curves (or even the type of curve which descends towards the upper limit K), the differences between the individual members depending on the initial state of the population and on the way in which the decrease in the inherent rate of increase takes place, whether through a decrease in fertility, or an increase in mortality, or a combination in varying degrees of both factors. Among the more interesting features of this type of population growth is the possibility, under suitable initial conditions, of the total numbers in the population becoming greater than K and then of finally approaching the stationary state by means of a series of damped oscillations around this limit.

It is interesting to consider in the light of these results some of the population growth curves which have been published for one or other species of insect living alone in a limited environment (e.g. Chapman, 1928; Crombie, 1945). Certainly the initial age distribution of some of these populations must have been extremely skew, consisting as they did in many cases of only a small number, perhaps only a pair, of adults. It is a little difficult, on looking through the figures given in these various papers, to rid oneself of the impression that some of the curves may have been influenced, in part at least, by these rather extreme initial conditions. But at present this remains an impression and nothing more; it does suggest, however, that the part played by the initial age distribution is worth investigating further in these experimental populations.

Although the dominant latent root of the matrix operating between t and $t+1$ has been considered here only as a function of the number of individuals present at time t, there should be little difficulty in extending the argument so as to include the case when q is assumed to be a function not of $N(t)$ but of $N(t-a)$ where a is an integer, or even of an integral, $\int_0^t N\,dt$ say. This last would be equivalent to assuming that the growth of the population was defined by a type of integro-differential equation such as is introduced by Volterra in his development of population mathematics (e.g. Volterra, 1931, p. 141; Volterra & D'Ancona, 1935, p. 22). Moreover, there is another and more speculative approach which is not without interest. In all these various forms of population growth the inherent rate of increase is regarded as dependent on the total numbers and thus each individual is counted as being of the same value for all age distributions of which it is a member. In other words, the factor q is taken to be some function of the scalar $[1]\xi$, where $[1]$ is a row vector of units. Now, from the biological point of view, it is not unreasonable to suppose that the form of the age distribution may also be of importance. For a given value of N we might have two entirely different age distributions, one of which was composed largely of adult individuals and only a small number of young, and the other with these proportions reversed. The question naturally arises whether one is justified in assuming that both the populations are of equal value and that they both influence the system of rates to the same extent. The one with the larger proportion of adults might exert a greater degree of influence on the rate of increase owing, for instance, to a proportionately greater consumption of food, or an enhanced mutual interference between the individual members of the population. But this is at present purely

speculative, and so far as the writer is aware, there is no experimental evidence for the occurrence of such differential effects associated with the form of the age distribution when the populations are of the same size. As a possibility, however, it is of interest theoretically and it suggests that instead of counting all individuals as equal, some system of weighting the individual age classes would be required. A mathematical model which immediately comes to mind is that of a matrix whose dominant latent root is affected by the length of the vector on which it is operating; that is to say, it would be assumed that the inherent rate of increase was dependent on the present value of the population at a given time.

6. The predator-prey relationship between two populations

It is of interest to consider very briefly a simple type of predator-prey relationship between two species of which the one, S_1, is preyed upon by the other, S_2. If the matrix A_1 with a dominant latent root λ_1 represents the optimum system of rates for the prey and the matrix A_{1t} for this population at time t has a dominant root λ_1/q_1, we might regard the factor q_1 as a function of N_2, the number of the predatory species S_2, and write as a first approximation,

$$q_1 = 1 + \alpha_1 N_2, \tag{6·1}$$

where $\alpha_1 > 0$ is a constant. In the same way there will be some optimum system of rates A_2 for the species S_2, though in fact this system may never be realized in full save under exceptional circumstances, for instance when the prey are extremely numerous in comparison with the predator, and everything in the environment is favourable to the latter species. (From the biological point of view there must be some upper limit to the possible inherent rate of increase of which a particular species is capable. For instance, in the case of mammals, this limit will be determined in part by physiological factors, such as the length of the gestation period, the shortest interval between litters, the maximum average number of daughters per litter, the age at which breeding first starts, and so forth, as well as the form of life table under the most favourable circumstances.) Then at time t the matrix A_{2t} will have a dominant root λ_2/q_2 and we will write

$$q_2 = 1 + \alpha_2 \frac{N_2}{N_1}, \tag{6·2}$$

where $\alpha_2 > 0$ is another constant and N_1 the number of the species S_1 at time t. This equation expresses in a simple fashion the main biological consequences to the species S_2 of its dependence upon S_1 as a source of food. For when $N_1 \to 0$, $q_2 \to \infty$, and the inherent rate of increase of the predator $r_2' = \log_e(\lambda_2/q_2) \to -\infty$ (disappearance of predator in the absence of any prey). Conversely, when N_1 becomes very large, $q_2 \to 1$ and the inherent rate of increase of the predator approaches its optimum value $r_2 = \log_e \lambda_2$.

Adopting, then, the simple system represented by (6·1) and (6·2) we shall have for the stationary state, putting $q_1 = \lambda_1$ and $q_2 = \lambda_2$,

$$N_1 = \frac{\alpha_2(\lambda_1 - 1)}{\alpha_1(\lambda_2 - 1)} = K_1, \quad N_2 = \frac{\lambda_1 - 1}{\alpha_1} = K_2,$$

which will be real positive quantities when both λ_1 and $\lambda_2 > 1$. Moreover, assuming for the moment that a stable stationary state is possible, we must have $\alpha_2(\lambda_1 - 1) > \alpha_1(\lambda_2 - 1)$ and

$(\lambda_1 - 1) > \alpha_1$ for both species to coexist in appreciable numbers. Then, expressing α_1 and α_2 in terms of the λ's and K's,

$$q_1 = 1 + (\lambda_1 - 1)\frac{N_2}{K_2}, \tag{6·1a}$$

$$q_2 = 1 + (\lambda_2 - 1)\frac{K_1 N_2}{K_2 N_1}. \tag{6·2a}$$

This simple system, however, can be improved upon to some extent. It will be noticed that if in equation (6·1) $N_2 = 0$, $q_1 = 1$ and thus in the absence of the predator it is assumed that the prey will increase to an unlimited extent. In order to introduce the conception of a limited environment, we might put

$$q_1 = 1 + \alpha_1 N_2 + \beta_1 N_1, \tag{6·3}$$

so that when $N_2 = 0$, the species S_1 will approach some upper limit in numbers. A slightly more general system is represented then by equations (6·3) and (6·2), for which the stationary state is

$$N_1 = \frac{\alpha_2(\lambda_1 - 1)}{\alpha_1(\lambda_2 - 1) + \alpha_2\beta_1}, \quad N_2 = \frac{(\lambda_1 - 1)(\lambda_2 - 1)}{\alpha_1(\lambda_2 - 1) + \alpha_2\beta_1}.$$

It would thus be possible to examine the consequences of various hypotheses as to the way in which the reduction in the optimum inherent rates of increase for the two species are effected. The possible combinations are, however, so numerous that it is difficult to cover at all adequately any more than one of the most obvious cases. In order to illustrate the properties of such a system, the simplest, and also the possibly not unrealistic example of the reduction in the rates for both species taking place through the operation of an additional force of mortality independent of age will be considered here. That is to say, it will be assumed that the effect of the species S_2 on system of rates for the species S_1 will be to divide the elements of the matrix A_1 by the factor q_1, and similarly that the effect of the species S_1 on the species S_2 and the matrix A_2 will be to divide the elements of the latter by q_2. This simplifies a number of the actual computations and also the analysis of the properties of the equations.

If at time t the age distributions of the $N_1(t)$ and $N_2(t)$ individuals of the species S_1 and S_2 are of the stable forms appropriate to the dominant latent roots λ_1 and λ_2 of the matrices A_1 and A_2 respectively, then from the properties of a matrix $q^{-1}A$ which were discussed in the previous section, the two populations will retain their initial forms of age distribution unchanged. The total numbers of individuals in the two populations, supposing these are subject to the system defined by equations (6·1a) and (6·2a) respectively, will therefore be at time $t+1$

$$N_1(t+1) = \frac{\lambda_1 N_1(t)}{1 + (\lambda_1 - 1)\dfrac{N_2(t)}{K_2}},$$

$$N_2(t+1) = \frac{\lambda_2 N_2(t)}{1 + (\lambda_2 - 1)\dfrac{K_1 N_2(t)}{K_2 N_1(t)}},$$

whence

$$N_1(t+1) - N_1(t) = \frac{(\lambda_1 - 1) N_1(t)\left\{1 - \dfrac{N_2(t)}{K_2}\right\}}{1 + (\lambda_1 - 1)\dfrac{N_2(t)}{K_2}},$$

and

$$N_2(t+1) - N_2(t) = \frac{(\lambda_2 - 1) N_2(t)\left\{1 - \dfrac{K_1 N_2(t)}{K_2 N_1(t)}\right\}}{1 + (\lambda_2 - 1)\dfrac{K_1 N_2(t)}{K_2 N_1(t)}}.$$

Before discussing the limits to which these difference equations will tend when the time interval is made smaller and smaller, it is necessary to consider the question of the value to which the dominant latent root of the matrix will tend when the latter becomes of a very large order. Suppose that working in some convenient unit of age and time we have the matrix A_1, with a real positive root λ_1, representing some given system of age-specific fertility and mortality rates. We can also construct a new matrix—$A_{\frac{1}{2}}$ say—for the same system of rates when the time interval is taken to be a half-unit. This new matrix will be twice the order of the original one and it will have a dominant root—$\lambda_{\frac{1}{2}}$ say—which will be less than λ_1. Continuing the process further, we shall have for an interval of age and time h a matrix A_h with a dominant root λ_h, this root representing in the case of a population with a stable age distribution, the ratio $N(t+h)/N(t)$. In order to compare the successive values of λ_h which would be obtained by making the interval h smaller and smaller, it is necessary to express them in some common unit of time and we can write

$$\Lambda = (\lambda_h)^{1/h} \quad \text{or} \quad \lambda_h = \Lambda^h.$$

Then, when the matrix remains constant in time, we shall have for a population with a stable age distribution,

$$\frac{N(t+h)-N(t)}{h} = \frac{\lambda_h - 1}{h} N(t) = \frac{\Lambda^h - 1}{h} N(t),$$

or, when $h \to 0$,

$$\frac{dN}{dt} = (\log_e \Lambda)\, N,$$

since

$$\lim_{h\to 0} \frac{\Lambda^h - 1}{h} = \log_e \Lambda.$$

Thus, as the matrix is made larger and larger, the value of $\log_e \Lambda$ tends to ρ, the true instantaneous relative rate of increase of the stable population per unit of time.

In a similar fashion we may write for an interval h the above difference equations in the form

$$\frac{N_1(t+h)-N_1(t)}{h} = \frac{\left(\dfrac{\Lambda_1^h - 1}{h}\right) N_1(t) \left\{1 - \dfrac{N_2(t)}{K_2}\right\}}{1 + (\Lambda_1^h - 1)\dfrac{N_2(t)}{K_2}},$$

$$\frac{N_2(t+h)-N_2(t)}{h} = \frac{\left(\dfrac{\Lambda_2^h - 1}{h}\right) N_2(t) \left\{1 - \dfrac{K_1 N_2(t)}{K_2 N_1(t)}\right\}}{1 + (\Lambda_2^h - 1)\dfrac{K_1 N_2(t)}{K_2 N_1(t)}},$$

which, as $h \to 0$, may be replaced by

$$\frac{dN_1}{dt} = (\log_e \Lambda_1)\, N_1 \left(1 - \frac{N_2}{K_2}\right), \quad \frac{dN_2}{dt} = (\log_e \Lambda_2)\, N_2 \left(1 - \frac{K_1 N_2}{K_2 N_1}\right).$$

Thus, when the age distributions of the populations S_1 and S_2 are each initially of the appropriate stable form, and when it is assumed that their respective systems of rates are represented by the matrices $q_1^{-1}A_1$ and $q_2^{-1}A_2$, the system of interrelations between the two populations which is defined by

$$q_1 = 1 + \alpha_1 N_2, \quad q_2 = 1 + \alpha_2 \frac{N_2}{N_1},$$

is equivalent to that defined by the differential equations

$$\frac{dN_1}{dt} = (r_1 - a_1 N_2) N_1, \quad \frac{dN_2}{dt} = \left(r_2 - a_2 \frac{N_2}{N_1}\right) N_2, \tag{6·4}$$

or, when q_1 is defined by (6·3), to

$$\frac{dN_1}{dt} = (r_1 - a_1 N_2 - b_1 N_1) N_1, \quad \frac{dN_2}{dt} = \left(r_2 - a_2 \frac{N_2}{N_1}\right) N_2, \tag{6·5}$$

where in both sets $r_1 = \log_e \lambda_1$, $r_2 = \log_e \lambda_2$ and a_1, a_2, b_1 are constants >0. This result is analogous to that discussed in the previous section for a single population increasing in a limited environment, where it was shown that when mortality was affected by a factor independent of age and the initial distribution was of the appropriate stable form, the numbers of individuals increased according to the logistic law, and that consequently under these conditions the type of population growth resulting from the operation of the matrix $q^{-1}A$, where

$$q = 1 + \frac{\lambda_1 - 1}{K} N,$$

was equivalent to that defined by the differential equation,

$$\frac{dN}{dt} = (r - aN) N.$$

The system of equations (6·4) differs somewhat from the classical Lotka-Volterra equations (Lotka, 1925, Chap. 8; Volterra, 1931, p. 14) for a simple predator-prey relationship between two species, in which the second member would be written

$$\frac{dN_2}{dt} = (-r_2 + a_2 N_1) N_2.$$

The form of the second member in (6·4) was originally suggested by the results of an analysis made by the author (unpublished observations) of some data given by Gause (1934) for the growth in numbers of *Paramecium caudatum* and *Paramecium aurelia* cultures, in which the food supply consisted of a suspension of *Bacillus pyocyaneus* in a buffered medium. Two different concentrations of bacteria—called by Gause 'one loop' and 'half-loop'—were used for both species of *Paramecium*, and under the conditions of the experiments these populations could be regarded as living in a limited environment with a constant supply of food. It was apparent from the results that for each species living alone the upper limit to the number of individuals depended on the concentration of food, being in each case approximately twice as great in the cultures with the 'one loop' concentration as in those with the 'half-loop'. If logistic equations are fitted to the four series of data given by Gause (1934, table 4, p. 145), it will be found that whereas the constant r in the equation $dN/dt = (r - aN) N$ remains approximately the same in the pair of experiments on each species of *Paramecium*, the constant a is inversely proportional to the concentration of food (see also on this point Kostitzin, 1937, p. 77). Thus, when the food supply (F) was kept constant, the form of population growth in numbers could be written

$$\frac{dF}{dt} = 0, \quad \frac{dN_2}{dt} = \left(r_2 - \frac{a_2}{C} N_2\right) N_2,$$

where C represents the relative concentration of food in the different experiments. This relationship suggested a system of equations such as (6·4) for the theoretical case of a food supply consisting of a population of individuals which when living alone would increase at

a rate $dN_1/dt = r_1 N_1$. However, apart from these considerations, the form of the second member of (6·4) is linked with the type of expression used here to define q_2 in terms of N_1 and N_2, and the latter arose as one of the simplest and most obvious ways of expressing the dependence of the species S_2 on S_1, bearing in mind that the elements of the matrix representing the system of rates at a given time must be positive quantities ($F_x \geqslant 0$, $0 < P_x \leqslant 1$). The difficulties which arise when this is not the case will be appreciated on endeavouring to find a working model in terms of matrices and vectors which will reduce to the classic Lotka-Volterra equations under suitable initial conditions. For, in the case of the predatory species S_2 we should have to consider a reciprocal matrix A_2^{-1} with a real positive root λ_2^{-1}, and at time t the matrix $q_2 A_2^{-1}$ would be regarded as operating on the vector $\xi(t)$ representing the age distribution of S_2. Then, if as before the matrix $q_1^{-1}A_1$ represents the system of rates for the species S_1 and

$$q_1 = 1 + \frac{\lambda_1 - 1}{K_2} N_2, \quad q_2 = 1 + \frac{\lambda_2 - 1}{K_1} N_1,$$

we have a system which will reduce to the Lotka-Volterra differential equations when the initial age distributions of both populations are of the stable form appropriate to their respective matrices A_1 and A_2. Now, apart from the fact that here no upper limit is placed on the inherent rate of increase, $r_2' = \log_e(q_2/\lambda_2)$, of the species S_2, there is an added complication that a number of the elements of A_2^{-1} will be negative (for the form of the matrix A^{-1} see the previous paper, § 4). Although no difficulties arise in the special case, when the age distribution of S_2 is of the stable form, in the perfectly general case of an arbitrary $\xi(t)$ some of the elements of $\xi(t+1) = q_2 A_2^{-1}\xi(t)$ can become negative and thus meaningless from the biological point of view. For these various reasons, therefore, the form of interrelationship between the two species defined by equations (6·1 a) and (6·2 a) was adopted here as a working model, and these reduce in the special case to the system of differential equations (6·4).

The writer has to confess that he has been unable to integrate either of the sets (6·4) and (6·5). Their main properties, however, seem to be quite clear. Taking the simplest system (6·4) first, we have for $dN_1/dt = dN_2/dt = 0$,

$$N_1 = \frac{r_1 a_2}{r_2 a_1} = K_1, \quad N_2 = \frac{r_1}{a_1} = K_2,$$

and, introducing for simplicity the variables $n_1 = N_1/K_1$, $n_2 = N_2/K_2$,

$$\frac{dn_1}{dt} = r_1 n_1(1 - n_2), \quad \frac{dn_2}{dt} = r_2 n_2\left(1 - \frac{n_2}{n_1}\right).$$

We will suppose that we are dealing with the case when $r_1 a_2 > r_2 a_1$, $r_1 > a_1$, in order that the stationary state may have a real meaning from the biological point of view. Then, in considering small departures from the stationary state, let $v_1 = n_1 - 1$, $v_2 = n_2 - 1$; and, disregarding in the usual way terms such as $v_1 v_2$, v_1^2, etc.;

$$\frac{dv_1}{dt} = -r_1 v_2, \quad \frac{dv_2}{dt} = r_2 v_1 - r_2 v_2.$$

This linear system will have a solution of the type $v_1 = A_1 e^{\mu_1 t} + B_1 e^{\mu_2 t}$, $v_2 = A_2 e^{\mu_1 t} + B_2 e^{\mu_2 t}$, where the values of μ will be given by the roots of the characteristic determinant

$$\begin{vmatrix} -\mu & -r_1 \\ r_2 & -(r_2 + \mu) \end{vmatrix} = 0$$

or

$$2\mu = -r_2 \pm \sqrt{(r_2^2 - 4r_1 r_2)}.$$

Thus, both roots μ_1 and μ_2 will be complex so long as $r_2 < 4r_1$, and the real part of this pair will be negative since $r_2 > 0$. The system under these conditions will therefore approach the stationary state by a series of damped oscillations. When $r_2 > 4r_1$, both μ_1 and μ_2 will be negative; and consequently the stationary state will be stable, since in both cases v_1 and v_2 tend to zero as time increases.

The analysis of the system represented by equations (6·5) leads to very similar results. For

$$\frac{dN_1}{dt} = \frac{dN_2}{dt} = 0,$$

$$N_1 = \frac{r_1 a_2}{r_2 a_1 + a_2 b_1} = K_1, \quad N_2 = \frac{r_1 r_2}{r_2 a_1 + a_2 b_1} = K_2.$$

And, in the same way as before, putting $n_1 = N_1/K_1$, $n_2 = N_2/K_2$, $v_1 = n_1 - 1$, $v_2 = n_2 - 1$, and neglecting terms in $v_1 v_2$, etc., we have

$$\frac{dv_1}{dt} = -r_1(1-k)\,v_1 - r_1 k v_2, \quad \frac{dv_2}{dt} = r_2 v_1 - r_2 v_2,$$

where

$$k = \left(1 + \frac{a_2 b_1}{r_2 a_1}\right)^{-1} \quad (0 < k < 1).$$

Then, putting the characteristic determinant

$$\begin{vmatrix} -\{r_1(1-k)+\mu\} & -r_1 k \\ r_2 & -(r_2+\mu) \end{vmatrix} = 0,$$

we have

$$\mu^2 + \{r_2 + r_1(1-k)\}\mu + r_1 r_2 k = 0.$$

The roots of this equation will be either both negative or both complex with the real part negative, depending on the relation between the various constants, and consequently both v_1 and v_2 will tend to zero as time goes on, the stationary state thus being stable as before. It will be noticed, however, that if $\mu = u \pm iv$, the damping term represented by the real part, $u = -\{r_2 + r_1(1-k)\}$, will be greater than in the case of the first system of equations (6·4) where $u = -r_2$. Again, for a given set of values of r_1, r_2, a_1, a_2, the number of individuals $N_1 = K_1$, $N_2 = K_2$ must be less for the second set of equations than for the first, since by definition $b_1 > 0$. Thus we might expect that for a population subject to equations (6·5) the stationary numbers will be lower and the approach to the stationary state more rapid than for a population subject to equations (6·4), provided that the values of r_1, r_2, a_1 and a_2 are the same in both cases.

As a numerical example of these predator-prey equations, suppose that the optimum system of rates for two imaginary species were the same and that they were represented by the matrix A which has been used previously to illustrate various points. Then $A_1 = A_2$, $\lambda_1 = \lambda_2 = 3$ and $r_1 = r_2 = 1{\cdot}09861$. If, for the first set of equations (6·1) and (6·2) we put

$$q_1 = 1 + 0{\cdot}002N_2, \tag{6·6}$$

$$q_2 = 1 + 10N_2/N_1 \tag{6·7}$$

and for the second, (6·3) and (6·2)

$$q_1 = 1 + 0{\cdot}002N_2 + 0{\cdot}0001851 85N_1, \tag{6·8}$$

q_2 remaining as before, the number of individuals for the stationary state are in the first case $K_1 = 5000$, $K_2 = 1000$, and in the second $K_1 = 3418$, $K_2 = 684$. Then, assuming that at $t = 0$, $N_1 = N_2 = 108$, and that each of these populations had the same stable form of age distribution

$$\xi_1 = \{81, 21, 5, 1\},$$

the results of operating on these two age distributions with the matrices $q_1^{-1}A_1$ and $q_2^{-1}A_2$ were as follows for the two sets of equations. The first two columns give the numbers of prey (N_1) and predators (N_2) when no upper limit is placed on the number of prey (equations (6·6) and (6·7)), and the second two columns give the respective numbers when the upper limit to N_1 would be 10800 individuals, if the predatory species was absent (equations (6·8) and (6·7)). (This is the same logistic population as was used in the previous section as an illustration.)

	I		II	
t	N_1	N_2	N_1	N_2
0	108	108	108	108
1	266	30	262	30
2	755	42	710	42
3	2089	81	1754	79
4	5393	175	3550	163
5	11983	396	5371	335
6	20050	894	6046	619
7	21583	1854	5403	917
8	13756	2991	4227	1020
9	5910	2826	3317	897
10	2665	1466	2921	726
11	2033	677	2928	625
12	2592	469	3146	598
13	4012	501	3396	618
14	6011	668	3555	658
15	7717	949	3587	692
16	7986	1277	3529	709
17	6741	1474		
18	5122	1388		
19	4071	1122		
20	3763	896		
21	4043	795		
22	4684	804		

In both cases the approach to the stationary state by means of a series of damped oscillations is very evident, this approach being made more rapidly in the second series than in the first as was to be expected from the results of the foregoing analysis. Probably the clearest graphical illustration of these functions is obtained by plotting $\log N_2$ against $\log N_1$, the result being a spiral curve which gradually approaches the stationary point.

Although these predator-prey equations have been studied here only in the special case of the reduction in the rates of increase of the two populations being effected by an increase in the degree of mortality which is independent of age, there would be little difficulty in investigating, for instance, the type of case in which a relative absence of prey affected the fertility of the predator, and so forth. Moreover, there will be in all cases the effect on such a system of any abnormalities in the initial age distributions, or of any chance disturbances of the existing age distributions at some point in the development of the populations. Without working out any actual examples, however, it might be expected from the results obtained in the case of a logistic-type population that the general effect of all these factors would be to add further oscillatory features to those which already are inherent in the system itself, even when the stability of the age-distributions is established as in the above numerical examples. It seems likely, too, that these additional factors will increase the chance of one or other of the two species being reduced to such low numbers as would be equivalent in practice to the extinction of the population. This possibility will, however, greatly depend on the numerical relations between the various constants which enter into the equations

and upon the initial conditions of the particular system. Finally, just as in the case of a solitary population increasing in a limited environment, there is the possibility of studying the more complicated cases in which q_1 and q_2 are taken to be functions not only of N_1 and N_2 at time t, but of the numbers at some previous time, or of an integral of N_1 or N_2 between some time limits. Similar methods could also be used in order to study a chain of such predator-prey relations.

This work arose out of some research carried out by the Bureau of Animal Population with the aid of a grant from the Agricultural Research Council, to which body grateful acknowledgement is made.

REFERENCES

CHAPMAN, R. N. (1928). The quantitative analysis of environmental factors. *Ecology*, **9**, 111–22.

CROMBIE, A. C. (1945). On competition between different species of graminivorous insects. *Proc. Roy Soc.* B, **132**, 362–95.

FISHER, R. A. (1930). *The Genetical Theory of Natural Selection.* Oxford: Clarendon Press.

GAUSE, G. F. (1934). *The Struggle for Existence.* Baltimore: Williams and Wilkins.

KOSTITZIN, V. A. (1937). *Biologie mathématique.* Paris: Armand Coli

LESLIE, P. H. (1945). On the use of matrices in certain population mathematics. *Biometrika*, **33**, 183–212.

LEWIS, E. G. (1942). On the generation and growth of a population. *Sankhya*, **6**, 93–6.

LOTKA, A. J. (1925). *Elements of Physical Biology.* Baltimore: Williams and Wilkins.

LOTKA, A. J. (1931). The structure of a growing population. *Hum. Biol.* **3**, 459–93.

LOTKA, A. J. (1939). Théorie analytique des associations biologiques. II. Analyse démographique avec application particulière à l'espèce humaine. *Actualités Sci.* no. 780, 1–149. Paris: Hermann.

VOLTERRA, V. (1931). *Leçons sur la théorie mathématique de la lutte pour la vie.* Paris: Gauthier-Villars.

VOLTERRA, V. & D'ANCONA, U. (1935). Les associations biologiques au point de vue mathématique. *Actualités Sci.* no. 243, 1–96. Paris: Hermann.

THE TRANSFORMATION OF POISSON, BINOMIAL AND NEGATIVE-BINOMIAL DATA

BY F. J. ANSCOMBE, *Rothamsted Experimental Station*

1. INTRODUCTION

Bartlett (1936) showed that if **r** is a Poisson variable with mean m and **y** is a random variable whose values y are derived by the transformation

$$y = \sqrt{r} \tag{1·1}$$

from the values r of **r**, then **y** is distributed rather more nearly normally than **r** with variance approximately $\frac{1}{4}$ if m is large, and the technique of analysis of variance may be applied to **y**.* He also showed that

$$y = \sqrt{(r+\tfrac{1}{2})} \tag{1·2}$$

is a better transformation, if slightly less convenient to use, as **y** then has more nearly a constant variance of $\frac{1}{4}$, even when m is not large. Similar transformations were proposed for a binomial variable, ar[illegible]ll (1942) gave the transformation analogous to (1·1) appropriate to a negative binon[illegible]le.

I begin by considering the transformation

$$y = \sqrt{(r+c)} \tag{1·3}$$

of a Poisson variable **r**, and show that for large m **y** has a most nearly constant variance (namely, $\frac{1}{4}$) when $c = \frac{3}{8}$; a result due to A. H. L. Johnson.

The similar transformation for a binomial variable **r**, with mean m and total number n, is

$$y = \sin^{-1}\sqrt{\left(\frac{r+c}{n+2c}\right)}. \tag{1·4}$$

The optimum value of c is $\frac{3}{8}$ if m and $n-m$ are large. The variance is approximately $\frac{1}{4}(n+\frac{1}{2})^{-1}$.

For a negative binomial variable **r**, with mean m and exponent k, the latter being constant and known, the corresponding transformation is

$$y = \sinh^{-1}\sqrt{\left(\frac{r+c}{k-2c}\right)}. \tag{1·5}$$

The optimum value of c is roughly $\frac{3}{8}$ if m is large and $k > 2$, and the variance is approximately $\frac{1}{4}\psi'(k)$, where $\psi'(t)$ denotes the second derivative of $\ln\Gamma(t)$ with respect to t. A simpler transformation, known to have an optimum property (i.e. to be the best of that degree of complexity) for m large and $k \geqslant 1$, is

$$y = \ln(r+\tfrac{1}{2}k); \tag{1·6}$$

the variance is approximately $\psi'(k)$. This is equivalent to setting $c = \frac{1}{2}k$ in (1·5). If k is large, $\psi'(k) \doteq 1/(k-\frac{1}{2})$ approximately.

The effect of these transformations for small values of m is shown numerically. The value $\frac{3}{8}$ for c appears to be nearly optimum for practical purposes, except with the negative binomial distribution when k is small. (When $k = 2$, the optimum value of c appears to be about 0·2.) In any case, it may be more convenient to choose a one-decimal value for c, i.e. generally 0·3 or 0·4.

* Letters in heavy type denote random variables, of which the same letter in light type denotes a possible value. Some equations, in particular all those of § 1, are equally valid for y and r in either type

,the 'angular transformation' given by
ne observed percentage as $100(r+c)/(n+2c)$,
e variance. For transformation (1·5), no corre-
published, I believe*.

POISSON DISTRIBUTION

. a Poisson distribution with mean m, and we consider the
, with c a non-negative constant.
$+c = m'$. Coefficients a_s are defined for $s = 1, 2, 3, \ldots$ by

$$a_s = (-1)^{s+1}\frac{1.(-1).(-3).\ldots.(-2s+3)}{2^s.s!}. \tag{2·1}$$

Then for any $t \geqslant -m'$ we have the Taylor series expansion

$$y = \sqrt{m'}\left\{1 + a_1\frac{t}{m'} - a_2\left(\frac{t}{m'}\right)^2 + \ldots + (-1)^s a_{s-1}\left(\frac{t}{m'}\right)^{s-1}\right\} + R_s. \tag{2·2}$$

If $t > 0$, we see at once from Lagrange's form of the remainder term that R_s satisfies

$$|R_s| < \frac{a_s t^s}{(m')^{s-\frac{1}{2}}}. \tag{2·3}$$

Considering now $|t| \leqslant m'$, we have directly from (2·2)

$$R_s(m')^{-\frac{1}{2}} = \left(1+\frac{t}{m'}\right)^{\frac{1}{2}} - \left\{1 + a_1\frac{t}{m'} - \ldots + (-1)^s a_{s-1}\left(\frac{t}{m'}\right)^{s-1}\right\} = \sum_{i=s}^{\infty}(-1)^{i+1}a_i\left(\frac{t}{m'}\right)^i. \tag{2·4}$$

The series converges, and we may write

$$\frac{R_s(m')^{s-\frac{1}{2}}}{t^s} = \sum_{i=s}^{\infty}(-1)^{i+1}a_i\left(\frac{t}{m'}\right)^{i-s}, \tag{2·5}$$

where the right-hand side again converges and is bounded. If $G(s)$ is a bound to its absolute magnitude,

$$|R_s| \leqslant G(s)\frac{|t|^s}{(m')^{s-\frac{1}{2}}}. \tag{2·6}$$

Comparing this inequality with (2·3), we see that it holds for all $t > -m'$.

We note now that the moments of **t** are

$$\mu_1 = 0, \quad \mu_2 = m, \quad \mu_3 = m, \quad \mu_4 = 3m^2 + m, \quad \text{etc.,} \tag{2·7}$$

and the absolute moment of order n is $O(m^{\frac{1}{2}n})$ as $m \to \infty$. We may therefore take expectations formally in the right-hand side of (2·2) and its powers, and derive asymptotic expansions for the moments of **y** as $m \to \infty$. We find

$$\operatorname{var}(\mathbf{y}) \sim \frac{1}{4}\left\{1 + \frac{3-8c}{8m} + \frac{32c^2 - 52c + 17}{32m^2}\right\}, \tag{2·8}$$

so that when $c = \frac{3}{8}$

$$\operatorname{var}(\mathbf{y}) \sim \frac{1}{4}\left\{1 + \frac{1}{16m^2}\right\}. \tag{2·9}$$

We have also

$$\mathrm{E}(\mathbf{y}) \sim \sqrt{(m+c)} - \frac{1}{8m^{\frac{1}{2}}} + \frac{24c-7}{128m^{\frac{3}{2}}}. \tag{2·10}$$

If we set

$$\mathrm{E}(\mathbf{y}) = \sqrt{(m_y + c)}, \tag{2·11}$$

* [Beall (1942, pp. 250–51) gave a table of $x^1 = k^{-\frac{1}{2}}\sinh^{-1}(kx)^{\frac{1}{2}}$, suited for the form of transformation which he used. Ed.]

m_y is the estimate of m derived by appl[illegible] metic mean $\bar{y}$ of a large sample of observe[illegible]

$$m_y \sim m - \frac{1}{4} + \frac{8[illegible]}{32m[illegible]}$$

so that setting $c = \frac{3}{8}$ also renders the bias $m_y - m$ in m_y [illegible]

Skewness and kurtosis of the distribution of $\mathbf{y}$ are mea[illegible]

$$\gamma_1 \sim -\frac{1}{2m^{\frac{1}{2}}}\left\{1 + \frac{25 - 48c}{16m}\right\},$$

$$\gamma_2 \sim \frac{1}{m}\left\{1 + \frac{945 - 1536c}{256m}\right\}. \tag{2·14}$$

These compare with $m^{-\frac{1}{2}}$ and m^{-1} for the original Poisson variable $\mathbf{r}$.

It is also of interest to find the large-sample efficiency E_y of the arithmetic mean $\bar{y}$ of observed values of $\mathbf{y}$ as a statistic for determining m (Bartlett, 1936, 1947). If $\mathbf{x}$ is a random variable having a distribution (absolutely continuous or discrete) such that the arithmetic mean $\bar{x}$ is a sufficient statistic for determining a parameter θ, it is easy to show, using the form of the frequency function of $\mathbf{x}$ given by Fisher (1934), that the large-sample efficiency E_y of the average $\bar{y}$ of any function $\mathbf{y}$ of $\mathbf{x}$, for determining θ, is the square of the correlation coefficient between x and y, i.e.

$$E_y = \frac{[\mathrm{cov}\,(\mathbf{x}, \mathbf{y})]^2}{\mathrm{var}\,(\mathbf{x})\,.\,\mathrm{var}\,(\mathbf{y})}. \tag{2·15}$$

In the present instance, $$E_y = \frac{[\mathrm{cov}\,(\mathbf{r}, \mathbf{y})]^2}{m\,\mathrm{var}\,(\mathbf{y})} \sim 1 - \frac{1}{8m} + \frac{16c - 9}{64m^2}. \tag{2·16}$$

3. Binomial distribution

We suppose now that $\mathbf{r}$ is distributed in a binomial distribution with total number n and mean m $(0 < m < n)$, and we consider the transformation

$$y = \sqrt{(n + d_2)}\sin^{-1}\sqrt{\left(\frac{r + c}{n + d_1}\right)}, \tag{3·1}$$

where c, d_1, d_2 are constants to be determined. Setting

$$r - m = t, \quad m + c = m', \quad n + d_1 = n_1, \quad n + d_2 = n_2,$$

the transformation becomes $$y = \sqrt{n_2}\sin^{-1}\sqrt{\left(\frac{m' + t}{n_1}\right)}. \tag{3·2}$$

We can expand y in a Taylor series in ascending powers of t, for $-m' \leqslant t \leqslant n_1 - m'$, and show that R_s, the remainder after s terms, is such that

$$\frac{n_1^s R_s}{t^s \sqrt{n_2}}$$

is bounded for all t in the range considered, with bounds depending on s and the ratio m'/n_1 only. Thus

$$|R_s| \leqslant G\left(s, \frac{m'}{n_1}\right)\frac{|t|^s}{n_1^s}\sqrt{n_2}. \tag{3·3}$$

he moments of y, valid for large n and

$$+\frac{3-8c}{8m}+\frac{3+8c-8d_1}{8(n-m)}\Big\},\tag{3·4}$$

$_2$, we have

$$+O\left(\frac{1}{n^2}\right).\tag{3·5}$$

$d_1 = 2c$, so that the transformation is symmetrical about $r = \frac{1}{2}n$. ffects the scale of y (for n fixed), and not the constancy of var (y) as e shape of the distribution of y.

quations corresponding to (2·12), (2·13), (2·14) and (2·16) are (proceeding to one term fewer in each case, for simplicity)

$$m_y \sim m+\frac{2m-n}{4n},\tag{3·6}$$

$$\gamma_1 \sim \frac{2m-n}{2\{nm(n-m)\}^{\frac{1}{2}}},\tag{3·7}$$

$$\gamma_2 \sim \frac{n^2-2m(n-m)}{nm(n-m)},\tag{3·8}$$

$$E_y \sim 1-\frac{(2m-n)^2}{8nm(n-m)}.\tag{3·9}$$

4. Negative binomial distribution

We can deal similarly with a negative binomial variable r, with mean m and exponent k ($m, k > 0$), such that the probability of observing a value r is

$$p_r = \frac{\Gamma(r+k)}{r!\,\Gamma(k)}\left(\frac{m}{m+k}\right)^r\left(1+\frac{m}{k}\right)^{-k} \quad (r = 0, 1, 2, \ldots).\tag{4·1}$$

The asymptotic expansions will be valid for large m and constant ratio k/m. We find the transformation

$$y = \sqrt{(k-\tfrac{1}{2})}\sinh^{-1}\sqrt{\left(\frac{r+\frac{3}{8}}{k-\frac{3}{4}}\right)},\tag{4·2}$$

with

$$\operatorname{var}(y) = \tfrac{1}{4}+O\left(\frac{1}{m^2}\right).\tag{4·3}$$

But it is of more interest to consider m large and k fixed. (The corresponding problem does not arise with the positive binomial, since $m < n$ necessarily.) The preceding method of obtaining the expansions now breaks down, as it relied on the ratio of standard deviation to mean of r tending to zero as $m \to \infty$. Now we have the ratio tending to $k^{-\frac{1}{2}}$.

We consider two transformations,

$$y = 2\sinh^{-1}\sqrt{\left(\frac{r+c}{k+d}\right)}\tag{4·4}$$

and

$$y = \ln(r+A).\tag{4·5}$$

It is supposed that c, $k+d$, and A are positive and constant. Apart from an added constant, (4·4) may be written

$$y = 2\ln\{\sqrt{(r+c)}+\sqrt{(r+c+k+d)}\}.\tag{4·6}$$

When r is large, we have (again ignorin

$$y \sim \ln$$

where for (4·4) or (4·6)

$$A = \tfrac{1}{2}(2c+k+d), \quad B^2 = \tfrac{1}{8}\{8c^2$$

and for (4·5)

$$B = A.$$

We proceed to find an asymptotic expansion (as $m \to \infty$
generating function of y, i.e. for

$$M(t) \equiv \sum_{r=0}^{\infty} e^{yt} p_r.$$

We set

$$\frac{m}{m+k} = e^{-\alpha}, \tag{2·14}$$

so that $\alpha \to 0$ as $m \to \infty$; and

$$e^{yt}\frac{\Gamma(r+k)}{\Gamma(k)\,r!}e^{-r\alpha} = u_r(\alpha). \tag{4·12}$$

We first prove the following

LEMMA. *As* $\alpha \to 0$,

$$\sum_{r=0}^{\infty} u_r(\alpha) - \int_0^{\infty} u_r(\alpha)\,dr \tag{4·13}$$

tends to a finite limit (depending on k and t, and on which function y of r is chosen, namely (4·4) *or* (4·5)).

Proof. By differentiating $\ln u_r(\alpha)$ with respect to r we see that as $r \to \infty$

$$\left(\frac{\partial}{\partial r}\right)^p u_r(\alpha) = u_r(\alpha)\left\{\frac{C}{r^p} + O\left(\frac{1}{r^{p+1}}\right)\right\} \quad (p = 1, 2, \ldots), \tag{4·14}$$

where C depends on $t+k-1$ and p but not on r or α, and the expression $O(r^{-(p+1)})$ is valid uniformly as $\alpha \to 0$. Since for large r

$$u_r(\alpha) = \frac{r^{t+k-1}e^{-r\alpha}}{\Gamma(k)}\left(1 + O\left(\frac{1}{r}\right)\right), \tag{4·15}$$

we see that

$$\left(\frac{\partial}{\partial r}\right)^p u_r(\alpha) = Ce^{-r\alpha} r^{t+k-1-p}\left(1 + O\left(\frac{1}{r}\right)\right) \tag{4·16}$$

as $r \to \infty$, uniformly as $\alpha \to 0$, where C satisfies the same condition as before. If $p > t+k+1$, we may write

$$\left|\left(\frac{\partial}{\partial r}\right)^p u_r(\alpha)\right| \le \frac{C}{(r+1)^2}, \tag{4·17}$$

where C is independent of α, and this holds for all $r \geqslant 0$ because the left-hand side is bounded for r in any bounded interval.

For all p, $(\partial/\partial r)^p u_r(\alpha) \to 0$ as $r \to \infty$, while at $r = 0$ $(\partial/\partial r)^p u_r(\alpha)$ is finite and tends to a finite limit as $\alpha \to 0$. We therefore have the Euler-Maclaurin expansion

$$\sum_{r=0}^{\infty} u_r(\alpha) = \int_0^{\infty} u_r(\alpha)\,dr + \tfrac{1}{2}u_0(\alpha) - \sum_{\nu=1}^{s-1}\frac{B_{2\nu}}{(2\nu)!}\left(\frac{\partial}{\partial r}\right)^{2\nu-1} u_0(\alpha) + R_s, \tag{4·18}$$

where

$$R_s = \int_0^{\infty} \theta(r)\left(\frac{\partial}{\partial r}\right)^{2s} u_r(\alpha)\,dr, \tag{4·19}$$

and $\theta(r)$ is a bounded periodic function of r, depending on s only. k is given, and we may restrict t to a neighbourhood of zero, say $|t| < \epsilon$. We choose s to be an integer satisfying $2s > k+1+\epsilon$. Then from (4·17) we see that R_s tends to a finite limit as $\alpha \to 0$, as do all the other terms after the first on the right-hand side of (4·18). Hence the lemma is proved.

$$(\alpha)\,dr + O(\alpha^k). \tag{4·20}$$

actly. We may, however, expand $u_r(\alpha)$ for
ng shown in (4·15) above. The error of the expansion
next term (independent of α) for $r \geqslant 1$. Integrating term
∞, we obtain after a tedious reduction the following

itions of (4·1), (4·7), (4·10) and (4·11), $M(t)$ can be expanded in the form

$$\frac{}{\Gamma(k)}\left[1 + (A - \tfrac{1}{2}k)\,t\frac{\alpha}{k+t-1}\right.$$
$$\left.+ \{(\tfrac{1}{2}(A - \tfrac{1}{2}k)^2 + \tfrac{1}{24}k)\,t^2 + (\tfrac{1}{2}kA - \tfrac{1}{24}k(k+3) - \tfrac{1}{2}B^2)\,t\}\frac{\alpha^2}{(k+t-1)(k+t-2)} + \dots\right] + O(\alpha^k). \tag{4·21}$$

The series in square brackets is continued as far as the term in α^n, where n is the greatest integer less than k. t is supposed confined to a neighbourhood of zero.

The cumulant-generating function is now found by taking the logarithm of $M(t)$, namely,

$$K(t) = -t\ln\alpha + \ln\Gamma(k+t) - \ln\Gamma(k) + (A - \tfrac{1}{2}k)\,t\frac{\alpha}{k+t-1} + \dots. \tag{4·22}$$

Hence
$$\operatorname{var}(\mathbf{y}) \sim \psi'(k) + \frac{k-2A}{(k-1)^2}\alpha, \tag{4·23}$$

if $k > 1$. We use the notation $\psi(t)$, $\psi'(t)$, etc., for the successive derivatives of $\ln\Gamma(k)$. If $A = \frac{1}{2}k$, we have

$$\operatorname{var}(\mathbf{y}) \sim \psi'(k) + \frac{k(k-1)(k-2) - (2k-3)(5k^2 - 3k - 12B^2)}{12(k-1)^2(k-2)^2}\alpha^2, \tag{4·24}$$

if $k > 2$. Considering $\mathbf{y}$ defined by (4·4), the condition $A = \frac{1}{2}k$ gives $d = -2c$, and the coefficient of α^2 in var (y) vanishes if c takes a value dependent on k, which for large k is approximately

$$c = \frac{3}{8} + \frac{23}{192}\frac{1}{k}, \tag{4·25}$$

and which rises to a little above 0·4 as k decreases towards 2. It appears from the numerical work below that a value of c somewhat under $\frac{3}{8}$ is optimum for practical purposes (if $k \geqslant 2$). For $\mathbf{y}$ defined by (4·5), we set $A = \frac{1}{2}k$ and then

$$\operatorname{var}(\mathbf{y}) \sim \psi'(k) - \frac{k(3k^2 - 9k + 7)}{12(k-1)^2(k-2)^2}\alpha^2, \tag{4·26}$$

if $k > 2$. If k is large and $m \gg k$, we have $\alpha \sim k/m$ and

$$\operatorname{var}(\mathbf{y}) \sim \frac{1}{k}\left(1 - \frac{k^2}{4m^2}\right). \tag{4·27}$$

Thus the larger k is the larger m must be for var (y) to approach its limiting value when $m \to \infty$. The transformation (4·5) is therefore not satisfactory if k is large.

For either form of transformation, we find the following limiting values as $m \to \infty$ $(\alpha \to 0)$:

$$m_y/m = \exp\{\psi(k) - \ln k\}, \tag{4·28}$$
$$\gamma_1 = \psi''(k)/[\psi'(k)]^{\frac{3}{2}}, \tag{4·29}$$
$$\gamma_2 = \psi'''(k)/[\psi'(k)]^2. \tag{4·30}$$

The effect of the transformation on the s
logarithmic transformation of a variance
Bartlett & Kendall (1946). We find also, by
of $\bar{y}$ in estimating m, namely, $E_y =$

For $\alpha > 0$, these expressions hold with error $O(\alpha^{*})$ v.

To conclude this investigation of asymptotic prope
(4·5) when $k = 1$. We note first that if $k < 2$, and t is suitab
(4·13) is also the limit as $N \to \infty$ of

$$\sum_{r=0}^{N} u_r(0) - \int_0^{N+\frac{1}{2}} u_r(0)\,dr.$$

When $k = 1$, $$u_r(0) = (r+A)^t,$$

and the limit is easily evaluated as (4·34)

$$f(A,t) = \tfrac{1}{2} + \ln\{\sqrt{(2\pi)}\,A^A e^{-A}/\Gamma(A)\}\,t + O(t^2).$$

We also note that the derivative of (4·13) with respect to α is finite and tends to a finite limit as $\alpha \to 0$ (proof as for the lemma itself), so that the error in replacing (4·13) by its limit as $\alpha \to 0$ is $O(\alpha)$. We thus find

$$M(t) = f(A,t)\,\alpha + (1-e^{-\alpha})\int_0^\infty (r+A)^t e^{-r\alpha}\,dr + O(\alpha^2). \tag{4·35}$$

The integral is an incomplete Γ-function and is easily expanded in powers of α. Hence

$$\begin{aligned}\operatorname{var}(\mathrm{y}) = {} & \psi'(1) - (A-\tfrac{1}{2})\,\alpha(\ln\alpha)^2 \\ & + 2\{(A-\tfrac{1}{2})\,\psi(1) + A - A\ln A + \ln(\sqrt{(2\pi)}\,A^A e^{-A}/\Gamma(A))\}\,\alpha\ln\alpha + O(\alpha).\end{aligned} \tag{4·36}$$

Setting $A = \frac{1}{2}$ (in agreement with our previous rule $A = \frac{1}{2}k$), and $\alpha = m^{-1}$, we have

$$\operatorname{var}(\mathrm{y}) = \psi'(1) - \ln 2\,\frac{\ln m}{m} + O\!\left(\frac{1}{m}\right). \tag{4·37}$$

5. Numerical investigation

Table 1 refers to the transformations (1·3), (1·4) and (1·5), with $c = \frac{3}{8}$ in each case. The first section gives the error of the estimate m_y of m got by applying the transformation in reverse to E (y). The second section gives the ratio of var (y) to the 'limiting variance'. The limiting variances are respectively $\frac{1}{4}$, $\frac{1}{4}(n+\frac{1}{2})^{-1}$ and $\frac{1}{4}\psi'(k)$; the first and third of these being the actual limits as $m \to \infty$, and the second the value suggested in § 3 for n large. The remaining sections of Table 1 give the shape constants γ_1 and γ_2 of the distribution of y and the efficiency E_y. With the binomial distribution, no entries are given for $m > \frac{1}{2}n$, as the functions are obviously symmetric or skew-symmetric about $m = \frac{1}{2}n$.

For the Poisson distribution, the variances may be compared with those given by Bartlett (1936) for $c = 0$ and $\frac{1}{2}$. For the negative binomial, Table 2 gives var (y) when $c = 0$ and $\frac{1}{2}$, and $k = 2$; and also var (y) for the transformation (1·6), when $k = 2$ and 5, the limiting variance being now $\psi'(k)$.

Comparing (1·5) and (1·6), we have indicated above that (1·5) is more effective in stabilizing the variance if $k > 2$, and the computations confirm this. If $c = \frac{3}{8}$, (1·5) is defined for all $k > \frac{3}{4}$, and it is quite possible that throughout this range it is superior to (1·6). As $k \to \frac{3}{4}$ (or, more generally, as $k \to 2c$) the two transformations become equivalent, and (1·6) is defined for all $k > 0$. Bearing in mind that (1·6) is more convenient to use than (1·5), it seems reasonable to recommend that (1·6) should be used if $k < 2$, and also for larger k if m is large; and otherwise (1·5).

tions with $c = \frac{3}{8}$

		(1·4)		Negative binomial (1·5)	
		$n = 10$	$n = 5$	$k = 5$	$k = 2$
		Bias in the mean $(m_y - m)$			
	9	−0·144	−0·107	−0·244	−0·330
	0·231	−0·142	−0·047	−0·395	−0·609
	−0·246	−0·100	—	−0·516	−0·869
	−0·250	−0·051	—	−0·625	−1·121
	−0·251	—	—	−0·830	−1·612
10	−0·250	—	—	−1·227	−2·577
20	−0·250	—	—	−2·209	−4·960
∞	−0·250	—	—	$-0{\cdot}098m$	$-0{\cdot}237m$
		Variance as fraction of the limiting variance			
1	0·717	0·759	0·802	0·639	0·537
2	0·924	0·958	0·990	0·848	0·731
3	0·983	1·001	—	0·929	0·824
4	0·999	1·006	—	0·964	0·875
6	1·002	—	—	0·990	0·927
10	1·001	—	—	0·999	0·966
20	1·000	—	—	1·000	0·989
∞	1·000	—	—	1·000	1·000
		Skewness coefficient γ_1			
1	0·31	0·30	0·31	0·34	0·41
2	−0·10	−0·07	0·01	−0·10	−0·04
3	−0·23	−0·13	—	−0·28	−0·25
4	−0·25	−0·08	—	−0·37	−0·37
6	−0·22	—	—	−0·45	−0·51
10	−0·17	—	—	−0·48	−0·63
20	−0·11	—	—	−0·48	−0·72
∞	0·00	—	—	−0·47	−0·78
		Kurtosis coefficient γ_2			
1	−0·80	−0·70	−0·54	−0·92	−1·02
2	−0·40	−0·21	0·06	−0·67	−0·93
3	−0·02	0·16	—	−0·32	−0·68
4	0·15	0·25	—	−0·06	−0·44
6	0·20	—	—	0·24	−0·08
10	0·12	—	—	0·43	0·35
20	0·05	—	—	0·46	0·77
∞	0·00	—	—	0·44	1·19
		Efficiency E_y			
1	96·5	97·7	98·9	93·7	89·9
2	95·8	97·9	99·4	91·3	85·5
3	96·1	98·9	—	90·2	83·0
4	96·7	99·5	—	89·9	81·7
6	97·8	—	—	89·8	80·2
10	98·7	—	—	89·9	78·9
20	99·4	—	—	90·2	78·0
∞	100·0	—	—	90·4	77·5

Table 2. *Other transformation*[illegible]

m	Transformation (1·[illegible] $k=2$			
	$c=0$	$c=\frac{1}{2}$	[illegible]	[illegible]
	Variance as fraction of the limi[illegible]			
1	1·106	0·503	0·456	
2	1·272	0·697	0·648	
3	1·255	0·793	0·749	
4	1·229	0·849	0·811	
6	1·174	0·908	0·879	0·8[illegible]
10	1·106	0·954	0·936	0·928
20	1·045	0·984	0·976	0·977
∞	1·000	1·000	1·000	1·000

The transformation (1·3), with $c = \frac{3}{8}$, derived by the formal expansion given in § 2, was communicated to me by Mr A. H. L. Johnson. He has kindly agreed to my publishing his result. I am indebted to Mr L. K. Turner and Dr A. A. Rayner for their patience in carrying out the laborious computations.

REFERENCES

Bartlett, M. S. (1936). The square root transformation in the analysis of variance. *J. R. Statist. Soc. Suppl.* **3**, 68.

Bartlett, M. S. (1947). The use of transformations. *Biometrics*, **3**, 39.

Bartlett, M. S. & Kendall, D. G. (1946). The statistical analysis of variance-heterogeneity and the logarithmic transformation. *J. R. Statist. Soc. Suppl.* **8**, 128.

Beall, G. (1942). The transformation of data from entomological field experiments so that the analysis of variance becomes applicable. *Biometrika*, **32**, 243.

Fisher, R. A. (1934). Two new properties of mathematical likelihood. *Proc. Roy. Soc.* A, **144**, 285.

Fisher, R. A. & Yates, F. (1938). *Statistical Tables for Biological, Agricultural and Medical Research* (3rd ed. 1948). Edinburgh: Oliver and Boyd.

SOME RESULTS IN THE TESTING OF SERIAL CORRELATION COEFFICIENTS

By M. H. QUENOUILLE

Anderson (1942) and Koopmans (1942) have recently investigated the distribution of the serial correlation coefficient defined by

$$r_l = \frac{\sum_{i=1}^{n} x_i x_{i+l}}{\sum_{i=1}^{n} x_i^2},$$

where the x_i are normally and independently distributed about zero, and x_{n+i} is taken as x_i. This definition approximates to the definition

$$r_l = \frac{\sum_{i=1}^{n-l} x_i x_{i+l}}{\sqrt{\left(\sum_{i=1}^{n-l} x_i^2 \sum_{i=1}^{n-l} x_{i+l}^2\right)}},$$

which is more generally used, and using this form, Anderson was able to obtain an exact distribution. Anderson's distribution was, however, difficult to use in practice, and Koopmans obtained an integral approximation, which was adequate for $n \geqslant 10$. Rubin (1945) has solved this integral to give the distribution

$$h(r)\,dr = \frac{\Gamma(\frac{1}{2}n+1)}{\Gamma(\frac{1}{2}n+\frac{1}{2})\,\Gamma(\frac{1}{2})}(1-r^2)^{\frac{1}{2}(n-1)}\,dr, \tag{1}$$

which is, in fact, the distribution of the ordinary correlation coefficient based on $n+3$ observations. Madow (1945) has demonstrated how this distribution may be extended, and if the x_i are connected by a linear Markoff scheme $x_{i+1} = \rho x_i + \epsilon_{i+1}$ where the ϵ_i are normally and independently distributed about zero, it is not difficult to show that a good approximation to the distribution of the first serial correlation coefficient is given by

$$h(r)\,dr = \frac{\Gamma(\frac{1}{2}n+1)}{\Gamma(\frac{1}{2}n+\frac{1}{2})\,\Gamma(\frac{1}{2})}\frac{(1-r^2)^{\frac{1}{2}(n-1)}}{(1-2\rho r+\rho^2)^{\frac{1}{2}n}}. \tag{2}$$

We can, in a manner analogous to the method used with the ordinary correlation coefficient, make the transformation

$$r = \tanh z, \quad \rho = \tanh\zeta \quad \text{and} \quad x = z-\zeta \text{ in the form (2).}$$

Then $$h(x)\,dx = \frac{\Gamma(\frac{1}{2}n+1)}{\Gamma(\frac{1}{2}n+\frac{1}{2})\,\Gamma(\frac{1}{2})}\frac{dx}{\cosh\zeta\cosh^{n+1}x(1+\tanh\zeta\tanh x)(1-\tanh^2\zeta\tanh^2 x)^{\frac{1}{2}n}}.$$

Now, for n large, $x = O(1/\sqrt{n})$ and

$$\log_e\frac{\Gamma(\frac{1}{2}n+1)}{\Gamma(\frac{1}{2}n+\frac{1}{2})\,\Gamma(\frac{1}{2})} = \tfrac{1}{2}\log_e\frac{n}{2\pi}+\frac{1}{4n}+O\left(\frac{1}{n^3}\right),$$

$$\log_e\cosh x = \frac{x^2}{2!}-\frac{2x^4}{4!}+O\left(\frac{1}{n^3}\right),$$

$$\log_e(1-\rho^2\tanh^2 x) = -\rho^2\left(x^2-\frac{2x^4}{3}\right)-\frac{\rho^4x^4}{2}+O\left(\frac{1}{n^3}\right),$$

$$\log_e(1+\rho\tanh x) = \rho\left(x-\frac{x^3}{3}\right)-\frac{\rho^2x^2}{2}+\frac{\rho^3x^3}{3}+O\left(\frac{1}{n^2}\right),$$

whence
$$h(x) = \frac{1}{\sqrt{(2\pi)}\,\sigma}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}X,$$
where
$$\sigma^2 = \frac{\cosh^2\zeta}{n} = \frac{1}{n(1-\rho^2)},$$
and
$$X = 1-\rho x+\left[\frac{1}{4n}-\frac{x^2(1-\rho^2)}{2}+\frac{nx^4(1-\rho^2)(1-3\rho^2)}{12}+\frac{\rho^2x^2}{2}\right]$$
$$-\rho x\left[\frac{1}{4n}-\frac{5x^2(1-\rho^2)}{6}+\frac{nx^4(1-\rho^2)(1-3\rho^2)}{12}+\frac{\rho^2x^2}{6}\right]+O\left(\frac{1}{n^2}\right).$$
We can now obtain the moments of the distribution in the form
$$\mu_1 = -\frac{\rho}{n(1-\rho^2)}+\frac{\rho(1+\rho^2)}{n^2(1-\rho^2)^2}+O\left(\frac{1}{n^3}\right),\quad \mu_2 = \frac{1}{n(1-\rho^2)}-\frac{2\rho^2}{n^2(1-\rho^2)^2}+O\left(\frac{1}{n^3}\right),$$
$$\mu_3 = \frac{6\rho^3}{n^3(1-\rho^2)^3}+O\left(\frac{1}{n^4}\right),\qquad \mu_4 = \frac{3}{n^2(1-\rho^2)^2}+\frac{2(1-9\rho^2)}{n^3(1-\rho^2)^3}+O\left(\frac{1}{n^4}\right),$$
and the coefficients of skewness and kurtosis are given by
$$\gamma_1 = \frac{6\rho^3}{[n(1-\rho^2)]^{\frac{3}{2}}}+O\left(\frac{1}{n^{\frac{5}{2}}}\right),\quad \gamma_2 = \frac{2(1-3\rho^2)}{n(1-\rho^2)}+O\left(\frac{1}{n^2}\right).$$
Hence, under this transformation, z will be distributed approximately normally about the mean
$$\zeta-\frac{\rho}{n(1-\rho^2)}+\frac{\rho(1+\rho^2)}{n^2(1-\rho^2)^2},$$
with variance
$$\frac{1}{n(1-\rho^2)}-\frac{2\rho^2}{n^2(1-\rho^2)^2}.$$
It should be noted that, compared with the transformation of the ordinary correlation coefficient, this transformation has the disadvantage that the variance of z is, to the first order, dependent upon the value of ρ. Furthermore, for $|\rho|$ large, the mean value of z deviates more widely from ζ.

The following examples are intended to investigate and demonstrate the testing of serial correlation coefficients when the error distribution of the ϵ is not normal.

In these examples, 1, 4 and 5 test the approximation by sampling artificial series with rectangular error distributions, 2 and 3 use an exceptional error distribution to show that the approximation is valid for n as low as twenty, while examples 6 and 7 demonstrate that the values of r obtained from certain observed series are consistent with homogeneity of the data.

Example 1. Serial correlation coefficients were calculated (according to the circular definition) from twenty sets of twenty random numbers, rectangularly distributed from -49 to 49.

The values of r thus obtained were:

-0.348,	-0.257,	-0.211,	-0.202,	-0.200,	-0.174,	-0.137,
-0.067,	-0.058,	-0.055,	-0.030,	0.011,	0.100,	0.112,
0.177,	0.234,	0.236,	0.348,	0.376,	0.445.	

These values were transformed by the transformation $r = \tanh z$ so that z should be distributed with zero mean, and variance 0·0500. The estimated mean and variance about zero were found to be 0·017 and 0·0530 respectively. The normality was, in addition, tested by the calculation of skewness and kurtosis coefficients
$$g_1 = 0{\cdot}595\pm 0{\cdot}512,\quad g_2 = -0{\cdot}751\pm 0{\cdot}992.$$

For small samples such as this, it must be borne in mind that the values of g_1 and g_2 do not demonstrate normality but only the absence of extreme non-normality.

Example 2. Suppose we consider the distribution $P(x_i = 1) = P(x_i = -1) = \frac{1}{2}$, so that $E(x_i x_{i+l}) = 0$. Then each product of successive observations will be distributed in the same manner as x_i, and r, based on n such products, which are independent, is distributed binomially (if a circular definition is employed, then the same distribution of r, omitting alternate ordinates, is obtained). Thus we can compare the terms of a binomial of degree n with the ordinates of the distribution of the ordinary correlation coefficient based on $n+3$ observations. The values of these are given below:

$n = 5$

r	Binomial terms	Correlation coefficient ordinates	Binomial terms ÷ 0·4
1·0	0·031	0·00	0·08
0·6	0·156	0·38	0·39
0·2	0·312	0·86	0·78

$n = 10$

r	Binomial terms	Correlation coefficient ordinates	Binomial terms ÷ 0·2
1·0	0·001	0·00	0·00
0·8	0·010	0·01	0·05
0·6	0·044	0·17	0·22
0·4	0·117	0·59	0·59
0·2	0·205	1·08	1·03
0·0	0·246	1·29	1·23

$n = 20$

r	Binomial terms	Correlation coefficient ordinates	Binomial terms ÷ 0·1
0·8	0·000	0·00	0·00
0·7	0·001	0·00	0·01
0·6	0·005	0·03	0·05
0·5	0·015	0·12	0·15
0·4	0·037	0·35	0·37
0·3	0·074	0·74	0·74
0·2	0·120	1·23	1·20
0·1	0·160	1·64	1·60
0·0	0·176	1·81	1·76

The comparison of the discrete and continuous distributions is difficult, but it is clear that, in view of the nature of the error distribution, the approximation is good for n as low as 20.

Example 3. The method of Example 2 can be used for the case when $\rho \neq 0$. If we consider the distribution

$$P(x_{i+1} = 1 | x_i = 1) = P(x_{i+1} = -1 | x_i = -1) = p,$$
$$P(x_{i+1} = 1 | x_i = -1) = P(x_{i+1} = -1 | x_i = 1) = q.$$

Then $E(x_i x_{i+l}) = (p-q)^l$, and this will behave similar to a linear Markoff scheme with $\rho = p-q$, except that only the values ± 1 are admissible. In this case, r is distributed binomially with parameter p so that we can compare the discrete binomial with the continuous theoretical distribution, as below. In this case, the fit is less satisfactory, since the binomial distribution is less skew, but, considering the approximate nature of the assumption, the fit is remarkable for n as low as 20.

$n = 20$

r	$p = \frac{2}{3}$ 10 (binomial terms)	$\rho = \frac{1}{3}$ Correlation coefficient ordinates	$p = \frac{3}{4}$ 10 (binomial terms)	$\rho = \frac{1}{2}$ Correlation coefficient ordinates
1·0	0·00	0·00	0·03	0·00
0·9	0·03	0·00	0·21	0·05
0·8	0·14	0·03	0·67	0·30
0·7	0·42	0·24	1·34	1·20
0·6	0·91	0·79	1·89	1·93
0·5	1·46	1·45	2·03	2·09
0·4	1·82	1·87	1·69	1·75
0·3	1·82	1·87	1·12	1·23
0·2	1·48	1·54	0·61	0·75
0·1	0·99	1·06	0·27	0·40
0·0	0·54	0·63	0·10	0·19
−0·1	0·25	0·32	0·03	0·08
−0·2	0·09	0·14	0·01	0·03
−0·3	0·03	0·05	0·00	0·01
−0·4	0·01	0·01	—	—
−0·5	0·00	0·00	—	—

Example 4. Serial correlation coefficients were calculated, using the ordinary definition, from thirty sets of twenty-one serial correlated numbers. These numbers had been derived using the scheme $x_{i+1} = 0{\cdot}5x_i + \epsilon_{i+1}$, where the ϵ_i were rectangularly distributed from -49 to 49. The thirty values of r thus obtained were

−0·113, 0·275, 0·282, 0·327, 0·339, 0·352, 0·386, 0·408, 0·411, 0·424,
0·428, 0·435, 0·445, 0·446, 0·476, 0·518, 0·547, 0·550, 0·562, 0·577,
0·590, 0·613, 0·614, 0·630, 0·645, 0·649, 0·672, 0·677, 0·718, 0·744.

These values were then transformed, and, in this case, we expect z to be distributed about mean

$$0{\cdot}5494 - \frac{0{\cdot}5}{20(0{\cdot}75)} + \frac{(0{\cdot}5)(1{\cdot}25)}{400(0{\cdot}75)^2} = 0{\cdot}5209,$$

with variance
$$\frac{1}{20(0{\cdot}75)} - \frac{0{\cdot}5}{400(0{\cdot}75)^2} = 0{\cdot}0644.$$

The estimated values of the mean and of the variance about the theoretical mean were 0·557 and 0·0517, while $g_1 = 0{\cdot}024 \pm 0{\cdot}427$ and $g_2 = 0{\cdot}785 \pm 0{\cdot}833$ indicated no significant deviation from normality.

Example 5. M. G. Kendall (1946) has calculated correlograms for eight sub-series of a series of 480 terms of the autoregressive scheme $x_{i+2} = 1\cdot1x_{i+1} - 0\cdot5x_i + \epsilon'_{i+2}$, where the ϵ'_i are rectangularly distributed from -49 to 49.

We might test his eight values of r_1 against the theoretical value $\rho_1 = 0\cdot733$. In this case, the ϵ_i in the scheme $x_{i+1} = \rho_1 x_i + \epsilon_{i+1}$ are correlated. The expected values of the mean and variance of the z distribution are easily found to be 0·9099 and 0·0353, and these may be compared with the mean, 1·0181, and the variance about the theoretical mean, 0·0454, calculated from the transformed values, 0·7498, 0·9181, 0·9223, 0·9439, 0·9962, 1·0352, 1·1786, 1·4007. The normality can again be tested giving

$$g_1 = 0\cdot567 \pm 0\cdot752, \quad g_2 = 0\cdot876 \pm 1\cdot481.$$

Example 6. Sir Gilbert Walker (1946) has calculated serial correlation coefficients for pressure data observed at daily intervals. His values for the period October to March in the years 1930–6 were 0·76, 0·88, 0·80, 0·88 0·79 and 0·82. These can be transformed and tested giving $s^2 = 0\cdot0215$ compared with the theoretical $\sigma^2 = 0\cdot0174$ approximately.

A similar set of unpublished data calculated for six successive months gave values of r_1 equal to 0·62, 0·76, 0·86, 0·78, 0·60, 0·86 with $s^2 = 0\cdot0688$ and $\sigma^2 = 0\cdot0803$ approximately.

Example 7. The Beveridge series of trend-free wheat-price index numbers was split into sixteen subseries of twenty-three terms each, and serial correlation coefficients were calculated for these subseries. The values obtained were 0·379, 0·460, 0·329, 0·421, 0·772, 0·464, 0·678, 0·767, 0·728, 0·311, 0·497, 0·656, 0·270, 0·508, 0·691 and 0·589 respectively. These were transformed, and the mean and variance of the transformed values were found to be 0·625 and 0·0625. If we assume $\rho = \tanh 0\cdot625 = 0\cdot5545$, then

$$\sigma^2 = \frac{1}{22(0\cdot6925)} - \frac{0\cdot6150}{484(0\cdot6925)^2} = 0\cdot0631,$$

which agrees remarkably well with the observed variance.

SUMMARY

We may summarize the results of these examples in tabular form, as below:

Example	ρ	σ^2	s^2	Degrees of freedom	χ^2	P
1	0·0	0·0500	0·0530	20	21·20	0·39
4	0·5	0·0644	0·0517	30	24·08	0·77
5	0·733	0·0353	0·0454	8	10·29	0·25
6 (*a*)	0·827	0·0174	0·0215	5	6·18	0·29
(*b*)	0·765	0·0803	0·0688	5	4·28	0·51
7	0·554	0·0631	0·0625	15	14·86	0·46
Total				83	80·89	0·55

Thus, while the results given here do not constitute definite proof, there is every indication that the approximate normal theory provides a satisfactory test for serial correlation coefficients, if the number of observations is sufficiently high, and such a test is undoubtedly

useful in numerous ways, varying from the determination of density of observations in sampling enquiries to the comparison of 'after-effects' in biological or economic research.

It is worth noting that the method of Examples 2 and 3 can be used to investigate the approximate test commonly used to test the correlation between two serially correlated series. If the correlation between successive terms of the two series are ρ_1 and ρ_2, then the correlation between corresponding terms of the series is usually tested with

$$n' = n(1-\rho_1\rho_2)/(1+\rho_1\rho_2)$$

degrees of freedom, i.e. as if the correlation were based on n' pairs of observations. This result was first proved by Bartlett (1935).

Suppose we consider two schemes of the kind described in Example 3, with parameters $\rho_1 = p_1 - q_1$ and $\rho_2 = p_2 - q_2$. Then if

$$p = p_1p_2 + q_1q_2 = \tfrac{1}{2}(1+\rho_1\rho_2), \quad q = p_1q_2 + p_2q_1 = \tfrac{1}{2}(1-\rho_1\rho_2) \quad \text{and} \quad \rho = p - q = \rho_1\rho_2,$$

it is not difficult to show that the distribution of the correlation r between the two schemes is given by

$$P(r=1) = P(r=-1) = p^{n-1},$$

$$P\left(r = 1-\frac{2i}{n}\right) = p^{n-2}q + (\tfrac{1}{2}n-1)\,p^{n-3}q^2 + [(\tfrac{1}{2}n-1)^2 - r^2(\tfrac{1}{2}n)^2]\,p^{n-4}q^3$$

$$+\frac{1}{1!\,2!}(\tfrac{1}{2}n-2)\,[(\tfrac{1}{2}n-1)^2 - r^2(\tfrac{1}{2}n)^2]\,p^{n-5}q^4$$

$$+\frac{1}{(2!)^2}[(\tfrac{1}{2}n-2)^2 - r^2(\tfrac{1}{2}n)^2]\,[(\tfrac{1}{2}n-1)^2 - r^2(\tfrac{1}{2}n)^2]\,p^{n-6}q^5 + \dots.$$

The ordinates of this distribution are given below, for $n = 20$ and

$$\rho = 1{\cdot}0,\ 0{\cdot}8,\ 0{\cdot}5,\ 0{\cdot}0,\ -0{\cdot}5,\ -0{\cdot}8,\ 1{\cdot}0,$$

$$p = 1{\cdot}0,\ 0{\cdot}9,\ 0{\cdot}75,\ 0{\cdot}5,\ 0{\cdot}25,\ 0{\cdot}1,\ 0{\cdot}0,$$

$$n' = 0,\ 2{\cdot}2,\ 6{\cdot}7,\ 20{\cdot}0,\ 60{\cdot}0,\ 180{\cdot}0,\ \infty.$$

r \ ρ	1·0	0·8	0·5	0	−0·5	−0·8	−1·0
1·0	0·5	0·0675	0·0021	—	—	—	—
0·9	0·0	0·0300	0·0056	—	—	—	—
0·8	0·0	0·0346	0·0118	0·0002	—	—	—
0·7	0·0	0·0389	0·0210	0·0011	—	—	—
0·6	0·0	0·0430	0·0332	0·0046	—	—	—
0·5	0·0	0·0466	0·0475	0·0148	0·0004	—	—
0·4	0·0	0·0498	0·0629	0·0370	0·0036	—	—
0·3	0·0	0·0523	0·0776	0·0739	0·0238	0·0013	—
0·2	0·0	0·0542	0·0898	0·1201	0·0954	0·0263	—
0·1	0·0	0·0553	0·0980	0·1602	0·2256	0·2219	—
0·0	0·0	0·0557	0·1008	0·1762	0·3025	0·5010	1·0000

A comparison of these values with tables of the distribution of the correlation coefficient constructed by David (1938) shows that they are distributed to a high degree of approxima-

tion with $n'+2$ degrees of freedom. The extra two degrees of freedom are more likely due to the nature of the error distribution than to the approximate nature of n', since they are independent of ρ. In any case, the conclusion reached is that the approximation is valid provided that the *effective* number of degrees of freedom is large.

REFERENCES

ANDERSON, R. L. (1942). *Ann. Math. Statist.* **13**, 1.

BARTLETT, M. S. (1935). *J. R. Statist. Soc.* **98**, 536.

DAVID, F. N. (1938). *Tables of the Ordinates and Probability Integral of the Distribution of the Correlation Coefficient in Small Samples.* Cambridge University Press.

KENDALL, M. G. (1946). *Contributions to the Study of Oscillatory Time Series.* National Institute of Economic and Social Research.

KOOPMANS, T. (1942). *Ann. Math. Statist.* **13**, 14.

MADOW, W. G. (1945). *Ann. Math. Statist.* **16**, 308.

RUBIN, H. (1945). *Ann. Math. Statist.* **16**, 211.

WALKER, Sir G. (1946). *Quart. J.R. Met. Soc.* **72**, 265.

FRACTIONAL REPLICATION ARRANGEMENTS FOR FACTORIAL EXPERIMENTS WITH FACTORS AT TWO LEVELS

By K. A. BROWNLEE, B. K. KELLY and P. K. LORAINE
The Research Department of the Distillers Company Ltd., Great Burgh, Epsom, Surrey

The theory of fractional replication of factorial experiments has been given recently (Finney, 1945; Finney, 1946; Kempthorne, 1947). The present paper is a description of the solutions of practical value in the case of the 2^{n-m} experiment (that is, the $(n-m)$th replicate of n factors all at two levels, involving the use of 2^m plots). Such confounding and double confounding arrangements as have been found will be given.

Two useful solutions are also included where one and two factors respectively are at four levels, these being derived from solutions with all factors at two levels.

It will be assumed that the reader is familiar with the basic concepts of fractional replication as set out by Finney (1945). These briefly are the group properties of treatment symbols and effect symbols, the use of generators to define concisely groups and subgroups, the concept of orthogonality between two subgroups, and the use of high-order interactions as aliases of main effects and first-order interactions.

In confounding with fractional replication the confounding subgroup with its aliases cannot contain any element in common with the alias subgroup except the identity and for practical purposes should not contain main effects nor, as far as possible, first-order interactions.

Fisher (1942a) introduced double confounding in which the two confounding subgroups must not contain any element other than the identity in common. In fractional replication there is the additional restriction that there should be no overlapping of the aliases.

To obtain the actual arrangement of plots in any particular simply confounded experiment, having selected the alias subgroup and the confounding subgroup, we form the so-called 'principal block' from those treatments whose symbols are orthogonal to both these subgroups. Subsequent blocks can be obtained by multiplying the principal block by symbols orthogonal to the alias subgroup which have not previously occurred. With double confounding these multipliers are of course the treatments of the principal block corresponding to the second confounding subgroup.

Enumeration of subgroups

Kempthorne (1947) reports that no simple method of enumerating subgroups suitable for high-order fractional replication has been found. It seems worth while, therefore, to record a method which has been found satisfactory for all designs of practical interest. It inevitably becomes more laborious for higher order subgroups, but such an enumeration is likely to be of sufficient use to justify the time spent in arriving at it.

Subgroups of order 2^m of the effects group involving n factors can be subdivided into types, two subgroups being considered of the same type when a permutation of the n letters representing the factors converts one into the other. Each type may be represented by a symbol

$$(n_1, n_2, n_3, \ldots),$$

where the (2^m-1) numbers n_1, n_2, etc., are the numbers of letters appearing in the separate elements of subgroups of that type. The symbolism may be further condensed by attaching to each number a suffix indicating how many times it is repeated. Thus the alias subgroup given below for the $\frac{1}{16}$ replicate of 8 factors would be represented by the symbol

$$(4_{14}, 8).$$

This symbolism is not perfect, as examples do occur where the same symbol represents more than one type. In such cases, the separate types will be distinguished by suffices outside the brackets.

The following conditions are *necessary* for a symbol to represent a type of subgroup:

(1) The sum of all the numbers in the brackets must equal 2^{m-1} times the number of letters actually used in the type.

(2) The numbers in the brackets must be all even or 2^{m-1} of them must be odd.

(3) When 2^{m-1} of the numbers are odd, the even numbers must, with the identity, form a subgroup of order 2^{m-1}.

(4) If the number n appears, the remaining numbers must be divisible into pairs such that the total of each pair is n.

(5) If the number 1 appears, the remaining numbers must be divisible into pairs such that the numbers in each pair differ by 1.

No other necessary conditions appear susceptible to expression in simple general terms.

The actual process of enumeration is best demonstrated by an example. Suppose it is required to find all the subgroups of order 8 involving 7 letters. The first step is to find all the combinations of 7, i.e. 2^m-1, even numbers which total to 28, i.e. $n \times 2^{m-1}$. Each of these is taken in turn and 2, i.e. 2^{m-2}, of the numbers are increased by 1 and the same number reduced by 1 in every possible way, bearing in mind the total number of elements containing a particular number of letters in the complete group. Next any duplication introduced by this process (some symbols will be derived from more than one combination of even numbers) is eliminated. The result will be all the symbols satisfying conditions (1) and (2) above. Further symbols are then rejected through failure to satisfy conditions (3), (4) and (5) above.

The final step consists in the examination of the remaining symbols for the existence of corresponding subgroups. This is best carried out by working from the even subgroup of order 2^{m-1} contained within it. Any one of the odd numbers may be taken for the last generator, and the number of factors this element must have in common with each of the even elements can be determined by the fact that the products must give the other odd elements. It can then be readily determined whether such a generator exists. The all-even symbols may be examined in a similar way except that there is a wider choice for the subgroup of order 2^{m-1}. At this stage, symbols representing more than one type of subgroup must be watched for. Considerable assistance can be obtained here by carrying out a check of the total number of subgroups.

The total number of subgroups of order 2^m in the group of order 2^n is given by (Carmichael, 1937)

$$\frac{(2^n-1)(2^n-2)(2^n-2^2)\dots(2^n-2^{m-1})}{(2^m-1)(2^m-2)(2^m-2^2)\dots(2^m-2^{m-1})}.$$

Some of these subgroups, however, do not use all the n factors. The number using all factors can be determined by a calculation exemplified in Table 1. The second column is derived by

applying the formula given by Carmichael. The third column is derived from the second column by subtracting $\binom{n}{3}$ from each entry. The fourth column is derived from the third by subtracting 11 times $\binom{n}{4}$ from each item and so on until the single entry in the sixth column is obtained. This gives the number of subgroups of order 8 which employ all the 7 factors, 7000 such subgroups employing 6 factors or less.

Table 1

(1) Factors n	(2) Total subgroups of order 8	(3)	(4)	(5)	(6)
3	1	—	—	—	—
4	15	11	—	—	—
5	155	145	90	—	—
6	1,395	1,375	1,210	670	—
7	11,811	11,776	11,391	9,501	4 811

When the existence of a subgroup corresponding to any symbol has been established, the number of subgroups of this type can be found by solving the appropriate combinatorial problem after one member of the type, or merely a set of generators, has been written out in full. Failure of the total to agree with that found as above after all symbols have been dealt with indicates, in the absence of errors, that at least one symbol represents more than one type of subgroup. In searching for this the all-even symbols should be examined first as these are usually, if not always, the source of the trouble.

If only subgroups containing no elements with less than five letters are considered of interest, the work can be considerably lightened by carrying out the above process but only starting from the all-even combinations which contain no numbers less than 4 and, in enumerating from these, only considering symbols in which all the 4's are raised to 5.

In this case, however, there does not seem to be any numerical check, so that there is some risk of missing one or more types through their being represented by the same symbols as other types. The risk is not very serious, however, since types represented by the same symbol have most properties which are important from the point of view of fractional replication in common.

Designs based on sixteen plots

There are designs based on 16 plots for 6, 7, or 8 factors ($\frac{1}{4}$, $\frac{1}{8}$, and $\frac{1}{16}$ replicates respectively), which give main effects with second-order interactions as aliases. First-order interactions must be used as error, so if these exist and are large the residual will be inflated accordingly.

The most general design is that for 8 factors. It is based on the alias subgroup of the type $(4_{14}, 8)$ generated by $ABCD,\ CDEF,\ ACEG,\ EFGH.$

This can be confounded in 4 blocks of 4, confounding AB, BC and AC, with

$$(1),\ abcd,\ efgh,\ abcdefgh,$$

as the first block and subsequent blocks being obtained by multiplying this block successively by *abef*, *aceg*, *adeh*.

The designs for 6 and 7 factors can be obtained from the above. The alias subgroups for these two cases are obtained by deleting those generators containing letters beyond the sixth and seventh respectively, and the plot treatments are obtained by deleting the letters beyond the sixth and seventh respectively where they occur.

DESIGNS BASED ON THIRTY-TWO PLOTS

The two extreme designs based on 32 plots are the half-replicate of 6 factors, which gives its main effects as aliases of fourth-order interactions and its first-order interactions as aliases of third-order interactions, and the $\frac{1}{32}$ replicate of 10 factors, which gives its main effects as aliases of second-order interactions and no first-order interactions. Between these two extremes come designs in which the order of aliases falls as the degree of completeness of replication falls.

The possibilities are set out in Table 2. The confounding arrangements for 6 and 7 factors can be used for double confounding, since in each case the two confounding subgroups and their aliases contain no element in common other than the identity. However, in the case of 7 factors, where it is desired merely simply to confound in 4 blocks of 8, it is better to use the confounding subgroup I, ACF, ACG, FG as FG is already lost, so nothing further is lost by this restriction.

It will be noted that the alias subgroup for the $\frac{1}{16}$ replicate has the structure $(4_7, 5_7, 9)$. At first sight one would have expected the subgroup generated by $ABCD$, $ABEF$, $ABGH$, $ACEGJ$ with the structure $(4_6, 5_8, 8)$ to be more satisfactory, but in point of fact, with this alias subgroup, only one factor retains its first-order interactions.

DESIGNS BASED ON SIXTY-FOUR PLOTS

The designs for 64 plots proceed from the half-replicate of 7 factors with a high degree of security in its aliases, first-order interactions having fourth-order interactions as aliases, to the $\frac{1}{32}$ replicate of 11 factors with all but three main effects having only second-order interactions as aliases.

The possibilities are shown in Table 3. With 8 blocks of 8, in all cases double confounding is possible with zero or slight loss of first-order interactions. With 16 blocks of 4 and 4 blocks of 16, satisfactory doubly confounded arrangements have been found only for 7, 8 and 11 factors, but simply confounded arrangements are given for 9 and 10 factors.

DESIGNS WITH MORE THAN SIXTY-FOUR PLOTS

For experiments with more than 64 plots the investment in most fields of experimentation is so great that it becomes essential to have a higher degree of certainty in the results, and hence the degree of ambiguity which can be tolerated with respect to the aliases is lower, i.e. the order of the interactions forming the aliases must be relatively higher.

The possibilities are set out in Table 4. The alias subgroup for 12 factors is due to Fisher (1942*b*). In all these designs the aliases of main effects are fourth-order interactions or better, and the aliases of first-order interactions are third-order interactions or better. The confounding arrangements given do not lose any first-order interactions.

There are certain less secure designs which may prove useful in certain circumstances and these are briefly summarized in Table 5.

Table 2. *Designs based on 32 plots*

No. of factors	6		7		8		9		10	
Order of replication	$\frac{1}{2}$		$\frac{1}{4}$		$\frac{1}{8}$		$\frac{1}{16}$		$\frac{1}{32}$	
Generators of alias subgroup	*ABCDEF*		*ABCDE* *CDFG*		*ABCD* *ABEF* *ACFGH*		*ABCD* *ABEF* *ACEG* *ABGHJ*		*ABCD* *CDEF* *EFGH* *GHJK* *ACEGJ*	
Main effects having as aliases interactions of order: 4	All		—		—		—		—	
3			*A, B, E*		*G, H*		*H, J*		—	
2			*C, D, F, G*		*A, B, C, D, E, F*		*A* to *G*		All	
First orders retained as third orders	All		0		0		0		0	
Factors retaining all first-order interactions as second orders	—		*A, B, E*		*G, H*		*H, J*		None	
No. of such interactions	—		15		13		15		0	
No. of other first orders with second orders as aliases	—		0		0		0		0	
No. of first orders lost	0		6		15		21		All	
Confounding in 4 blocks of 8 and 8 blocks of 4:	Double				Single					
Generators of confounding subgroups	*ABC* *AEF*	*ACE* *BDE* *BCF*	*CG* *BCD*	*ACF* *FG* *BG*	*CD* *DF*	*CD* *DF* *AD*	*AG* *EG*	*AG* *EG* *AF*	*CD* *DF*	*CD* *DF* *BC*
Further first-order interactions lost other than those contained in the confounding subgroup	*AD*	*AB* *CD* *EF*	*AE*	*CE* *DE*	—	—	—	—	—	—
Total no. of new first-order interactions lost by confounding	1	3	1	4	0	0	0	0	0	0
Generators of principal blocks	*ef* *bc* *acde*	*abcd* *abef*	*ae* *bcg* *bdf*	*acde* *abfg*	*gh* *abg* *cdef*	*gh* *abcdefg*	*hj* *aceg* *bdfh*	*hj* *abcdefgh*	*cdef* *abgh* *ghjk*	*ghjk* *abcdefgh*

Table 3. *Designs based on 64 plots*

No. of factors	7		8		9		10		11	
Order of replication	$\frac{1}{2}$		$\frac{1}{4}$		$\frac{1}{8}$		$\frac{1}{16}$		$\frac{1}{32}$	
Generators of alias subgroup	*ABCDEFG*		*ABCDE* *ABFGH*		*ABCD* *ABEFG* *ACEHJ*		*ADHJ* *BEGK* *ABCGH* *ABFJK*		*ABCD* *CDEF* *EFGH* *GHJKL* *ACEGJ*	
Main effects having as aliases interactions of order: 5	All		—		—		—		—	
4	—		—		—		—		—	
3	—		All		*E, F, G, H, J*		*C, F*		*J, K, L*	
2	—		—		*A, B, C, D*		*A, B, D, E, G, H, J, K*		*A* to *H*	
No. of first orders retained as fourth orders	All		0		0		0		0	
No. of first orders retained as third orders	—		9		0		0		0	
Factors retaining all first-order interactions as second orders	—		*A, B*		*E* to *J*		*C, F*		*J, K, L*	
No. of such interactions	—		13		30		17		27	
No. of other first orders with second orders as aliases	—		6		0		16		0	
No. of first orders completely lost	0		0		6		12		28	
Double confounding in 8 blocks of 8:										
Generators of confounding subgroups	*ABC* *CDE* *BDF*	*ABD* *DEF* *BFG*	*AB* *ADF* *ACG*	*AH* *ADG* *BEF*	*BC* *CD* *BE*	*FJ* *AH* *HJ*	*AB* *BC* *AJ*	*AF* *AG* *BE*	*AB* *BC* *AJ*	*AF* *JK* *AG*
Further first-order interactions lost other than those contained in the confounding subgroups	—	—	*EH*	*EG* *BC*	*AE*	*AG*	—	*CD* *FK*	*DJ*	—
Total no. of new first-order interactions lost by confounding	0	0	2	3	4	7	5	5	4	1
Generators of principal blocks	*acdf* *bcef* *bcdg*	*bcdf* *adfg* *abef*	*cegh* *abcd* *cdfg*	*defg* *bcf* *adh*	*abcdefh* *fghj* *hj*	*bce* *bdg* *acfhj*	*degh* *efk* *abcdhjk*	*cdh* *behj* *afgjk*	*efgh* *kl* *abcdghjl*	*abefghjk* *bcehl* *cdjk*
	Double				Single				Double	
Confounding in 16 blocks of 4 and 4 blocks of 16:										
Generators of confounding subgroups	*AB* *BCF* *CDG* *DE*	*ABCD* *ABEG*	*AB* *BCF* *CDG* *DEH*	*CDF* *ADEH*	*AE* *FJ* *AH* *BC*	*BC* *CD*	*AB* *BC* *BE* *BJ*	*AD* *DH*	*AF* *JK* *AG* *AK*	*AB* *BC*
Further first-order interactions lost other than those contained in the confounding subgroups	*AG* *BG* *EF* *DF*	—	*AE* *BE* *EG* *BG* *DF* *AG* *CH*	—	*BE* *DE* *DG* *DH* *EG* *GH*	—	*DF*	—	*CL*	—
Total no. of new first-order interactions lost by confounding	6	0	8	0	10	0	9	0	8	0
Generators of principal blocks	*cdef* *abcg*	*ab* *adef* *eg* *cd*	*abdefg* *cdfh*	*defg* *abcd* *abgh* *bcf*	*abcdegh* *bcfj*	*abcd* *hj* *fg* *efh*.	*abcegjk* *dfghk*	*bcef* *adhj* *bck* *bfg*	*bcehl* *abefghjk*	*abcd* *ghjk* *efgh* *kl*

Table 4. *Designs with more than* 64 *plots*

No. of factors	9	12	16
Order of replication	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{1}{128}$
No. of plots	128	256	512
Generators of alias subgroup	*ABCDEF* *DEFGHJ* — — — — —	*ABEFLM* *ABGHJK* *ACEGKM* *ADFGJM* — — —	*ABCDEFGH* *EFGHJKLM* *ABEFJKNO* *ACEGJLPQ* *ADEHJMNP* *ADEGNQ* *ACJMNO*
Confounding in 8 blocks: Generators of confounding subgroup	*ADG* *BEH* *CFJ*	*ABE* *CDE* *ACG*	*ACE* *BDE* *ADF*
Generators of principal block	*acdf* *abde* *degh* *efhj* — —	*abcd* *bcegjk* *acehlm* *jklm* *fhjl* —	*abcd* *jklm* *nopq* *adegjk* *bdfhkmn* *ghjloq*
Multipliers by which subsequent blocks are obtained	*ab* *bc* *ac* *cfj* *cfg* *cfh* *cdh*	*bceh* *afk* *bfj* *bek* *aej* *adjm* *efjk*	*acehp* *bcefn* *abfhnp* *cgmpq* *aeghmq* *befgmo* *dfghmnq*

Table 5

No. of factors	10	11	13
Order of replication	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$
No. of plots	128	128	256
Generators of alias subgroup	*ABCDG* *ABEFH* *AGHJK* — —	*ADEKL* *AFGHJ* *BDFJL* *CDGJK* —	*ABCDEF* *DEFGHJ* *GHJKLM* *ADGKN* *BEHKM*
Order of alias of main effects	Third	Third	Third
No. of first-order interactions:			
With third orders as aliases	17	7	34
With second orders as aliases	28	48	44

In the case of the $2^{10/7}$, for confounding in 4 blocks of 32, the confounding subgroup can be

$$I = CEJK = ABGH = ABCEGHJK,$$

and for 8 blocks of 16 the further generator $DFGH$ can be added, when the two first-order interactions AG and CE are lost.

EXPERIMENTS WITH SOME FACTORS AT FOUR LEVELS

A factor at four levels can be regarded as two factors at two levels, the third degree of freedom coming from their interaction. In adopting solutions obtained for factors all at two levels, this interaction must be regarded as a main effect from the point of view of its occurrence in the sets of aliases. This makes the problem of obtaining satisfactory arrangements more difficult, and here only two solutions will be given.

(a) *The half-replicate of one factor at four levels and four factors at two levels*

We can use $I = ABCDEF$ as the alias subgroup, $I = ABC = ADE = BCDE$ as the confounding subgroup for 4 blocks of 8, and allocate B and D to represent the four-level factor. This loses AF, and leads to the arrangement in Table 6, where the first number in each treatment symbol represents the four-level factors and the remainder the factors A, C, E and F.

Table 6

10000	21011	00111	31100
10011	21000	00100	31111
20111	11100	30000	01011
01101	30110	11010	20001
01110	30101	11001	20010
20100	11111	30011	01000
31010	00001	21101	10110
31001	00010	21110	10101

(b) *The half-replicate of two factors at four levels and three factors at two levels*

We use the alias subgroup $I = ABCDEFG$, the confounding subgroup

$$I = AEG = CFG = ACEF,$$

and allocate A and B to the first four-level factor and C and D to the second. All the degrees of freedom for the main effects have third-order interactions as aliases, and all the degrees of freedom for the first-order interactions have second-order interactions as aliases. The resulting arrangement is in Table 7.

Table 7

11000	31000	13000	22000
01111	21111	03111	32111
32001	12001	30001	01001
22110	02110	20110	11110
00000	20000	02000	33000
10111	30111	12111	23111
23001	03001	21001	10001
33110	13110	31110	00110
21100	01100	23100	12100
30100	10100	32100	03100
20011	00011	22011	13011
31011	11011	33011	02011
12010	32010	10010	21010
03010	23010	01010	30010
13101	33101	11101	20101
02101	22101	00101	31101

Acknowledgements are due to the Directors of the Distillers Company Limited for permission to publish this paper.

REFERENCES

CARMICHAEL, R. D. (1937). *Introduction to the Theory of Groups of Finite Order*, §28. Boston: Ginn.

FINNEY, D. J. (1945). The fractional replication of factorial arrangements. *Ann. Eugen., Lond.*, **12**, 291–301.

FINNEY, D. J. (1946). Recent developments in the design of field experiments. III. Fractional replication. *J. Agric. Sci.* **36**, 184–91.

FISHER, R. A. (1942*a*). *The Design of Experiments*, 3rd ed., §45.2. London: Oliver and Boyd.

FISHER, R. A. (1942*b*). The theory of confounding in factorial experiments in relation to the theory of groups. *Ann. Eugen., Lond.*, **11**, 341–53.

KEMPTHORNE, O. (1947). A simple approach to confounding and fractional replication in factorial experiments. *Biometrika*, **34**, 255–72.

THE RELATIONSHIP BETWEEN FINITE GROUPS AND COMPLETELY ORTHOGONAL SQUARES, CUBES, AND HYPER-CUBES

By K. A. BROWNLEE and P. K. LORAINE

The Research Department of the Distillers Company, Ltd., Epsom, Surrey

Completely orthogonal squares of side n afford a means of introducing $(n+1)$ factors all at n levels into an experiment involving n^2 plots instead of the n^{n+1} that would be required for a full factorial experiment. It is pertinent to inquire how far interactions between pairs of factors will affect the estimates of the main effects of the other factors. The theory of fractional replication, as based on the theory of finite prime power groups by Finney (1945), allows an approach to be made to this problem. Completely orthogonal hyper-cubes will also be investigated.

Finney has indicated the theory of fractional replication of experiments involving n factors each at π levels, where π is any prime. The treatment combinations can be represented by symbols $a^\alpha b^\beta c^\gamma \ldots$, where the indices $\alpha, \beta, \gamma, \ldots$ take only the values $0, 1, 2, \ldots (\pi-1)$. If the symbols are interpreted according to the ordinary laws of algebra, with the additional condition

$$a^\pi = b^\pi = c^\pi = \ldots = 1,$$

then it is found that the product of any two symbols is a third. The complete set of symbols for a full factorial experiment form a prime power group of modulus π and order π^n.

Two symbols $a^\alpha b^\beta c^\gamma \ldots$ and $a^{\alpha'} b^{\beta'} c^{\gamma'} \ldots$ are said to be orthogonal if

$$\alpha\alpha' + \beta\beta' + \gamma\gamma' + \ldots = 0 \pmod{\pi}.$$

If we have a subgroup of order π^p it is possible to select a second subgroup of order π^{n-p} such that all its elements are orthogonal to those of the first subgroup; this second subgroup is said to be the complete orthogonal subgroup of the first subgroup. Finney has shown that if the first subgroup represents the treatments carried out in a fractional replication of a factorial experiment then the effects are interconfused in sets of $(n-p)$, and to find the aliases of any effect we multiply the second subgroup by that effect. The second subgroup is therefore called the alias subgroup.

Formation of completely orthogonal squares of any prime number

To construct a completely orthogonal square of side π, where π is a prime, we need to allocate to each of the π^2 plots their row and column numbers and their levels of the remaining $(\pi-1)$ factors. The square can be considered either to define the levels of $(\pi+1)$ factors or to define the letters of $(\pi-1)$ alphabets in a square.

Consider the treatment subgroup of π^2 elements generated by $x_1 P_1$, $x_2 P_2$, where

$$P_1 = abc \ldots w, \quad P_2 = ab^2c^3 \ldots w^{\pi-1},$$

and w is the $(\pi-1)$th alphabet. The exponents of x_1 and x_2, reduced to modulus π, will run from 0 to $(\pi-1)$ and can therefore be used to define the cells in the square with co-ordinates y_1, y_2. The element corresponding to cell (y_1, y_2) will be $(x_1 P_1)^{y_1} (x_2 P_2)^{y_2}$ and the levels of the $(\pi-1)$ alphabets are the exponents, reduced to modulus π, of the letters a to w in this product.

It is obvious that every level of each alphabet occurs just once in each row and each column. To see that this square is completely orthogonal, we need also to show that each level of any alphabet occurs once and once only with each level of any other alphabet. Suppose that the levels of the rth and sth alphabets ($r \neq s$) are the same for two different cells (p_1, p_2) and (q_1, q_2). Then if their exponents in P_i are ρ_i and σ_i respectively ($i = 1, 2$), we have

$$\rho_1 p_1 + \rho_2 p_2 = \rho_1 q_1 + \rho_2 q_2, \quad \sigma_1 p_1 + \sigma_2 p_2 = \sigma_1 q_1 + \sigma_2 q_2$$

and since $p_1 \neq q_1$, $p_2 \neq q_2$, this leads to $$\frac{\rho_1}{\rho_2} = \frac{\sigma_1}{\sigma_2}. \tag{1}$$

Here $$\rho_i = 1 + (i-1)(r-1), \quad \sigma_i = 1 + (i-1)(s-1), \quad (i = 1, 2),$$

which gives the contradiction $r = s$. Hence the square is completely orthogonal.

The present treatment using prime power groups is essentially that of Stevens (1939). We can proceed to derive the alias subgroup of order $\pi^{\pi-1}$ completely orthogonal to the above treatment subgroup of order π^2. It can be generated by $X_1 X_2 A^{\pi-1}$ and the $(\pi-2)$ other elements of the form $X_1^t A^{\pi-t-1}\Omega_{t+1}$, where $t = 1, 2, \ldots, \pi-2$ and Ω_u is the uth alphabet. For example, in the case of $\pi = 5$ the treatment subgroup generators are $x_1 abcd$, $x_2 ab^2c^3d^4$, and the alias subgroup generators are $X_1X_2A^4$, X_1A^3B, $X_1^2A^2C$, X_1^3AD.

Squares of sides 4 and 8

The present treatment using prime-power groups is not immediately adaptable to the case of squares of sides of the form π^n, but we can deal with the case of 2^2 and 2^3 in the following manner.

Fisher & Yates (1943) give as the treatments for the completely orthogonal square of side 4, writing in order columns, rows and the three alphabets,

11111, 12234, 13342, 14423, 21222, etc.

Let us represent each four-level factor by two two-level factors, so that A, B and AB correspond to the 3 degrees of freedom for the first four-level factor, etc. Thus

$$1 = b, \quad 2 = (1), \quad 3 = a \quad \text{and} \quad 4 = ab.$$

The treatment combinations then can be written

bdfhk, *bgjk*, *bcegh*, *bcdefj*, *d*, etc.

This particular set of elements does not form a subgroup, as it does not contain the identity. It can be converted into a subgroup, however, by multiplying through by any of its elements. Selecting d as the multiplier, we get

bfhk, *bdgjh*, *bcdegh*, *bcefg*, (1), etc.

This subgroup can be generated by

bfhk, *bdgjh*, *cehjk*, *aegj*.

We have 5 four-level factors or 10 two-level factors, so the full factorial experiment will have $4^5 = 2^{10}$ elements and the alias subgroup completely orthogonal to the treatment subgroup will have $2^{10-4} = 2^6$ elements and can be generated by

$$BDF, \quad ACE, \quad BEGH, \quad AEFH, \quad AFJK, \quad DHK.$$

The treatment of squares of side 8 is similar to those of side 4. We represent the first 8-level factor by the three 2-level factors A, B and C, etc., such that

$$1 = (1), \quad 2 = a, \quad 3 = b, \quad 4 = ab, \quad 5 = c, \quad 6 = ac, \quad 7 = bc, \quad 8 = abc.$$

The treatment subgroup of $8^2 = 2^6$ elements can be generated by

dgmopqrstuwyα, *ehkprstuvwxzβ*, *fjlmnopqrtvxz*, *agknqtwz*, *bhloruxα*, *cjmpovyβ*.

The complete group corresponding to the full factorial experiment will have $8^9 = 2^{27}$ elements, so the alias subgroup completely orthogonal to the above treatment subgroup will have $2^{27-6} = 2^{21}$ elements.

The selection of 21 elements from the 2,097,152 to act as generators is not too easy, but a systematic method of approach leads to

$$CDFN,\ EMP,\ DNQ,\ DHV,\ DTZ,\ BFL,\ FGHJKLM,$$
$$FPS,\ BGNO,\ CMSV,\ FTN,\ DRX,\ CHOP,\ VF\beta,$$
$$AHQR,\ NWX\alpha;\ FMY,\ JNPT,\ ADG,\ BEH,\ CFJ.$$

SQUARES WITH LESS THAN THE COMPLETE NUMBER OF FACTORS

As they represent the most general condition, we have discussed the completely orthogonal squares. The above treatments can, however, be immediately adapted to squares with less than the complete number of factors.

To consider a Latin square of side 5, for example, the complete treatment group for the full factorial experiment has 5^3 elements. The treatment subgroup as before has 5^2 elements, and these can be obtained from those previously given by missing out the last three letters, i.e. the generators are $x_1 a$, $x_2 a$. The alias subgroup will now have $5^{3-2} = 5$ elements and may be generated from $X_1 X_2 A^4$.

For a Graeco-Latin square the alias subgroup will have $5^{4-2} = 5^2$ elements, and we add the second generator given previously, $X_1 A^3 B$. Similarly, for a hyper-Graeco-Latin square the alias subgroup has $5^{5-2} = 5^3$ elements, and we add the third generator given previously, $X_1^2 A^2 C$.

In all cases above for each factor omitted the appropriate treatment subgroup is obtained by missing out from the generators the letters corresponding to the omitted factor, and the appropriate alias subgroup is obtained by omitting those generators containing letters corresponding to the omitted factor.

RELATIONSHIP BETWEEN MAIN EFFECTS AND INTERACTIONS BETWEEN OTHER FACTORS

In all cases for squares ranging from Latin to completely orthogonal the alias subgroups contain three-letter elements of the type $A^\alpha B^\beta C^\gamma$, and every factor occurs in such elements. Thus all the degrees of freedom for the A main effect have as aliases terms of the type $B^{\beta'} C^{\gamma'}$, i.e. first-order interactions. In using orthogonal squares for experimental designs it is therefore essential that no pair of factors should be liable to interact. In view of the unsatisfactory nature of this result, one is led to inquire whether Latin or completely orthogonal cubes or hyper-cubes will provide a sounder basis for experimental designs.

ORTHOGONALITY IN THREE DIMENSIONS

Consider a treatment subgroup generated by elements of the form $x_1 P_1$, $x_2 P_2$, $x_3 P_3$, where

$$P_1 = abcd \ldots w, \quad P_2 = ab^2c^3d^4 \ldots w^{\pi-1}, \quad P_3 = ab^3c^5d^7 \ldots w^{2\pi-3},$$

in which w is the $(\pi-1)$th letter, and whenever the exponent of one of the letters a to w becomes 0 (mod π) such a letter is deleted from all the generators. The letter to be deleted will in fact be the $\left(\frac{\pi+1}{2}\right)$th. This treatment subgroup of π^3 elements can be considered to define the levels of $(\pi+1)$ factors or the levels of $(\pi-2)$ alphabets in a cube, the position

co-ordinates in this cube being given by the x_i. The alphabets are numbered before the deletion is made and for convenience this numbering will be retained. The levels of the alphabets in cell (y_1, y_2, y_3) are given by the exponents of the letters a to w in the element $(x_1 P_1)^{y_1} (x_2 P_2)^{y_2} (x_3 P_3)^{y_3}$.

The condition for such a cube to be Latin is that in any 'file' defined by two of the y_i being constant, each of the π levels of each alphabet should occur once and once only. This condition is obviously fulfilled here.

For the cube to be completely orthogonal we must also have that each plane section $y_i = \text{const.}$ $(i = 1, 2, 3)$ forms an orthogonal square. The arguments used above to obtain (1) hold here if suffices j and k are substituted for 1 and 2 throughout $(j, k = 1, 2, 3)$. Therefore we have

$$\frac{\rho_j}{\rho_k} = \frac{\sigma_j}{\sigma_k},$$

where $\rho_i = 1 + (i-1)(r-1)$, $\sigma_i = 1 + (i-1)(s-1)$, $(i = j, k)$, which leads to $r = s$ as before.

The alias subgroup will be of order $\pi^{\pi-2}$ and has $(\pi - 2)$ generators of the form

$$X_1 X_2 X_3 A^{\pi-1}, \quad X_1^t X_3^{\pi-t} A^{\pi-t-1} \Omega_{t+1},$$

where $t = 1, 2, \ldots, \frac{\pi-1}{2}, \frac{\pi+3}{2}, \ldots, \pi - 2$, and Ω_u is the uth alphabet.

Orthogonality in four or more dimensions

Considerations analogous to those for three dimensions show that for four dimensions the treatment subgroup generators will be

$$x_1 abc \ldots w, \quad x_2 ab^2c^3 \ldots w^{\pi-1}, \quad x_3 ab^3c^5 \ldots w^{2\pi-3}, \quad x_4 ab^4c^7 \ldots w^{3\pi-5},$$

where, as before, whenever the exponent of one of the letters a to w becomes $0 \pmod \pi$, such a letter is deleted from all generators. Thus the $\left(\frac{\pi+1}{2}\right)$th letter will be deleted and also the $\left(\frac{\pi+2}{3}\right)$th or the $\left(\frac{2\pi+2}{3}\right)$th, according as π is of the form $(3k+1)$ or $(3k+2)$, k being an integer.

The generalization to n dimensions is now obvious. The treatment subgroup will have n generators of the form $x_i P_i$, where

$$P_i = ab^{1+(i-1)} c^{1+2(i-1)} \ldots w^{1+(\pi-2)(i-1)} \quad (i = 1, 2, \ldots, n),$$

deletions being made as before. After deletions there will thus always be a total of $(\pi + 1)$ classifications or factors.

The proofs of complete orthogonality follow exactly as for three dimensions.

The treatment subgroup is of order π^n and therefore the alias subgroup is of order $\pi^{\pi+1-n}$ and has generators of the form

$$X_1 X_2 \ldots X_n A^{\pi-1},$$
$$X_1^{\pi+t} X_2^{\pi+0.t} X_3^{\pi-1.t} X_4^{\pi-2.t} \ldots X_n^{\pi-(n-3)t} A^{\pi-t-1} \Omega_{t+1},$$

where Ω_u is the uth alphabet and, if $m_1, m_2, \ldots, m_{n-2}$ are the numbers in ascending order of the omitted alphabets, then

$$t = 1, 2, \ldots, (m_1 - 1), (m_1 + 1), \ldots, (m_2 - 1), (m_2 + 1), \ldots, (m_{n-2} - 1), (m_{n-2} + 1), \ldots, \pi - 2.$$

EXPERIMENTAL DESIGNS: LATIN HYPER-CUBES

Inspection of the generators for the treatment subgroups shows that a Latin (i.e. having only one alphabet) hyper-cube in any number of dimensions is always possible for any prime. For example, we can have a Latin hyper-cube in four dimensions for $\pi = 3$. This will have 5 classifications and be based on 3^4 cells; it is thus a solution of the one-third replicate of 5 factors at three levels. This solution is not actually the same as given by Finney (1945) as he selected one more satisfactory for confounding. Similarly the one-fifth replicate of 6 factors at five levels can be generated immediately and its alias subgroup will have the one generator $X_1X_2X_3X_4X_5A^4$ thus allowing first-order interactions to be estimated with third orders as aliases.

EXPERIMENTAL DESIGNS: GRAECO-LATIN HYPER-CUBES

In view of the form of the generators for the treatment subgroup, and in particular the property of elimination of letters where exponents become 0 (mod π), it is clearly impossible to generate, with this method, n-dimensional hyper-cubes other than Latin except when $\pi \geqslant n+1$.

In the case of Graeco-Latin hyper-cubes we have two alphabets and n co-ordinates making up $(2+n)$ factors in π^n plots. The order of replication is therefore π^{-2}, and the alias subgroup, generated by

$$X_1X_2 \dots X_n A^{\pi-1}, \quad X_1X_3^{\pi-1} \dots X_n^{\pi-(n-2)} A^{\pi-2} B,$$

will have no element with less than $(n+1)$ letters.

In the case of three dimensions, for $\pi = 3$ only a Latin cube is possible. With $\pi \geqslant 5$, the Graeco-Latin is possible, and the main effects of all factors will have second-order interactions as aliases. For example, with $\pi = 5$ we can have 5 factors with $5^3 = 125$ plots.

In the case of four dimensions, π must be $\geqslant 5$. For example, with $\pi = 5$, the two alphabets and four co-ordinates make 6 factors in $5^4 = 625$ plots. Main effects have third-order interactions and first-order interactions have second-order interactions as their lowest order aliases.

On going to five dimensions the minimum prime is 7 and hence the minimum number of plots is $7^5 = 16807$.

EXPERIMENTAL DESIGNS: COMPLETELY ORTHOGONAL HYPER-CUBES

Consideration of the general alias subgroup shows that if this is of order π^m, $m \geqslant 3$, then its generators must include $(m-1)$ of the type

$$X_1^t X_3^{\pi-t} \dots X_n^{\pi-(n-2)t} A^{\pi-t-1} \Omega_{t+1}.$$

Since the subgroup generated by any pair of elements of this type contains elements of the form $A^\alpha B^\beta C^\gamma$, and since every alphabet occurs in such elements, all the factors corresponding to the alphabets include first-order interactions amongst their aliases. The factors corresponding to the co-ordinates will be better, however, having as aliases interactions of order $(n-1)$.

For example, with $\pi = 5$, in three dimensions, we can have three alphabets. The total number of plots is $5^3 = 125$ and the order of replication $5^{-3} = 1/125$. Three factors have first-order aliases and three factors have second-order aliases. With $\pi = 7$, the total number of plots is $7^3 = 343$, being a $7^{-5} = 1/16807$ replicate. Five factors have first-order interactions as aliases and three have second-order aliases.

ADDITIONAL NOTE ON 'THE DESIGN OF OPTIMUM MULTIFACTORIAL EXPERIMENTS'

Kempthorne (1947) has pointed out that the multifactorial designs of Plackett & Burman (1946) are in some cases high-order fractionally replicated factorial designs.

Thus their case of $N = 16$, $L = 2$, $n = 15$ is a 2^{-11} replicate of the 2^{15} experiment. The alias subgroup of 2^{11} elements for their design can be generated by

$$ABN, \quad ACK, \quad ADE, \quad AFL, \quad AGJ, \quad AHO, \quad AMP, \quad BCO, \quad BDL, \quad BGM, \quad CDP.$$

Altogether there are 35 three-letter elements in the alias subgroup, and each factor has, amongst its aliases, 7 first-order interactions involving the other 14 factors.

Their case of $N = 32$, $L = 2$, $n = 31$ has an alias subgroup of $2^{31-5} = 2^{26}$ elements containing 155 three-letter elements. Thus each factor has, amongst its aliases, 15 first-order interactions involving the other 30 factors.

The three cases where $N = 9$, 25 and 49, $L = 3$, 5 and 7, and $n = 4$, 6 and 8 respectively are completely orthogonal squares with the alias characteristics described above.

With $N = 27$, $L = 3$, $n = 13$, if we ascribe the letters a to n (omitting i) to the 13 factors in the order given, the treatments are a subgroup generated by

$$ad^2f^2gh^2j^2km^2n^2, \quad bde^2fgh^2klm, \quad cef^2ghj^2lmn,$$

and the alias subgroup of 3^{10} elements is generated by

$$BCL^2, \quad A^2B^2K, \quad D^2EG^2, \quad EGN, \quad CEL, \quad C^2K^2G, \quad B^2J^2F, \quad ACJ, \quad AL^2M, \quad C^2HK.$$

Thus in all the multifactorial designs amenable to representation as fractional replications, all main effects contain first-order interactions amongst their aliases.

Thanks are due to the Distillers Company Limited for permission to publish this paper.

REFERENCES

FINNEY, D. J. (1945). The fractional replication of factorial arrangements. *Ann. Eugen., Lond.*, **12**, 291–301.

FISHER, R. A. & YATES, F. (1948). *Statistical Tables for Biological, Agricultural, and Medical Research* (3rd edition). London: Oliver and Boyd.

KEMPTHORNE, O. (1947). A simple approach to confounding and fractional replication in factorial experiments. *Biometrika*, **34**, 255–72.

PLACKETT, R. L. & BURMAN, J. P. (1946). The design of optimum multifactorial experiments. *Biometrika*, **33**, 305–25.

STEVENS, W. L. (1939). The completely orthogonalized Latin square. *Ann. Eugen., Lond.*, **9**, 82–93.

SYSTEMATIC SAMPLING OF CONTINUOUS PARAMETER POPULATIONS

A. E. JONES

[*Editorial note.* This paper was submitted by Dr Jones in the autumn of 1947. A number of points arising on the original draft were discussed with him, and the Editors were about to consult him on further amendments which they thought desirable when Dr Jones met an untimely death in a lift accident on 7 May 1948. The amendments have, nevertheless, been made in the interests of clarity, and the more important of them are indicated in this paper. No alteration has been made in the substance of the paper or in the main formulae, but the second part has been omitted and is discussed in the succeeding paper by Mr Kendall.]

INTRODUCTION

1. A sample is usually said to be distributed systematically when the individual members are selected according to some deterministic rule. For instance, if it were desired to estimate the mean of a characteristic of a mass-produced article, and a sample consisting of every tenth article were taken, we might say that this was a *systematic* one. Such a sample may, of course, be random if the method of selection of the samples is unconnected, in any conceivable manner, with the value of the characteristics under consideration.

Estimates obtained from a systematic sample may sometimes be more accurate than those from a random sample of the same size, but it is usually considered that this lower accuracy is more than counterbalanced by the fact that a more reliable estimate of the sampling variation can be obtained by the latter method. The customary formulae giving an estimate of sampling error are based on the assumption that there is no correlation between the individual members of the sample. It sometimes happens that neighbouring sample members are positively correlated, in which case the error estimated from the customary formulae may differ substantially from the true value.

The problem to be considered here is that of estimating by a systematic sample the 'average value' of a random variable from a one-dimensional homogeneous population depending on a continuous parameter, such as the nitrogen content of soil along a one-dimensional strip. The phrase 'average value' requires closer definition. Suppose $Z(t)$ be used to denote the actual nitrogen content at a fixed point distance t from one end of the strip under consideration. Then $Z(t)$ may be regarded as a random variable with a hypothetical population consisting of all the nitrogen contents at this point throughout a number of years. There is an infinity of random variables corresponding to all the points of the t-interval considered. We are not interested in the means of these populations, but in the value of the nitrogen content of the whole of the strip of soil. Thus, if T_0 and T ($> T_0$) are the ends of the interval, the average value in which we are interested is

$$\bar{Z} = \frac{1}{T - T_0}\int_{T_0}^{T} Z(t)\,dt,$$

which is not a parameter of any of the hypothetical populations mentioned above, and, in fact, is itself a random variable.

It will only be possible to make a finite number of observations and if, say, n is the maximum number it is proposed to make, we would wish to dispose them in such a way that

the mean-square error of our estimate of $\bar{Z}$ will be a minimum. Hence if $z(t_1), z(t_2), \ldots, z(t_n)$ represent observed variates at $t_1, t_2, \ldots, t_n$ respectively, and if our best* estimate of $\bar{Z}$ is the linear combination

$$a_1 z(t_1) + a_2 z(t_2) + \ldots + a_n z(t_n),$$

the mean-square error of the estimate will be

$$S = E[a_1 z(t_1) + a_2 z(t_2) + \ldots + a_n z(t_n) - \bar{Z}]^2, \tag{1·1}$$

where the operator E denotes expected value.

Apart from determining $t_1, t_2, \ldots, t_n$ we also require the mean-square error of this estimate of $\bar{Z}$. Mathematically, this resolves into two problems: first, that of obtaining the mean-square error as a function of the (unknown) parameters of the distribution of $Z(t)$, and secondly obtaining a formula in terms of actual sample values which will provide a satisfactory estimate of this function.

It will be shown, on reasonable hypotheses, that if the correlation between the successive values at the points $t_1, t_2, \ldots, t_n$ is very small, the best method of distributing n members along a strip of length T is at distances $T/(n+1)$, $2T/(n+1)$, ... from one end, while if the correlation is greater than 0·25 the best distribution will be at distances $T/2n$, $3T/2n$, ...

Theoretical representation of $Z(t)$—assumptions

2. It is necessary to make certain assumptions about the way in which the random function $Z(t)$ varies. Physically, we should expect that there will be some smooth curve, or trend, about which $Z(t)$ will fluctuate randomly. In addition, we should expect that $Z(t)$ will be continuous in t, in other words we should expect that by making h small we could make $[Z(t+h) - Z(t)]$ as small as we please. Incidentally, it follows from this that it is impossible to regard the random variables $Z(t)$ and $Z(t+h)$ as independent for all values of h.

Thus any recorded value, $z(t_1)$, is the sum of three components: (*a*) an error of observation; (*b*) the value of the trend at t_1; (*c*) a random variable, which will be denoted by $X(t_1)$, representing $Z(t_1)$ less the value of the trend at t_1. From the definition of the trend we may take

$$EX(t_1) = 0 \quad \text{for all } t_1. \tag{2·1}$$

The following assumptions will be made about these three components:

(A) That the error of observation, i.e. $z(t_1) - Z(t_1)$, is a random variable with zero mean, and is uncorrelated with all other errors of observation, and with $Z(t)$ for all t. $x(t_1)$ will be used to denote the sum of $X(t_1)$ and the error of observation at t_1. In other words, $x(t_1)$ is $z(t_1)$ corrected for trend.

(B) In the first place it will be assumed that the trend is simply a constant for all t. If

$$\bar{X} = \frac{1}{T - T_0}\int_{T_0}^{T} X(t)\,dt, \tag{2·2}$$

the problem of minimizing S of equation (1·1) then reduces to the problem of minimizing

$$E\{a_1 x(t_1) + a_2 x(t_2) + \ldots + a_n x(t_n) - \bar{X}\}^2, \tag{2·3}$$

* Throughout this paper the expressions 'best estimate' or 'optimum estimate' will be used to mean the linear estimate whose mean-square error is a minimum. [The expression 'mean-square error' is used instead of 'variance' because we are considering a minimization of the type $E(X-Y)^2$, where X and Y are random variables, not of the type $E\{X - E(x)\}^2$.—Ed.]

provided we confine our attention to unbiased estimates, i.e. make the restriction*

$$\sum_{j=1}^{n} a_j = 1. \tag{2·4}$$

(C) About $X(t)$ it will be assumed that:

(1) The variance of $X(t) = E\{X(t)\}^2 = \lambda$, a constant independent of t. (2·5)

(2) The correlation of $X(s)$ and $X(t)$ depends only on $|s-t|$, and decreases exponentially as $|s-t|$ increases. $E\{X(s)X(s+t)\}$ will be denoted by $\lambda R(t)$, where $R(t)$ is the correlation function of $X(t)$. Thus

$$R(t) = e^{-q|t|} \quad (q>0). \tag{2·6}$$

It follows that for h small

$$E\{X(t+h)-X(t)\}^2 = 2\lambda(1-e^{-q|h|}) \tag{2·7}$$

is also small, and since

$$E\{X(t+h)-X(t)\} = 0,$$

it will be seen that by these assumptions we imply that $X(t)$ is continuous in the mean for all t.†

I consider that the above assumptions are the simplest that satisfy the required physical conditions. (A) represents nothing new. Although it is assumed that the trend is a constant for all t, the results are valid if the trend is a linear function of t. As regards the assumption (C) it is worth noting, though it will not be discussed here, that there is a 'small-error' justification for (2·5) and (2·6) similar to the usual small-error justification for the assumption of normality. They are both derived from the Central Limit Theorem.

I have made some practical attempts to decide whether the above assumptions are satisfactory for a number of meteorological variables, e.g. wind speed and air temperature. It was found that the observations analysed could be satisfactorily interpreted on these hypotheses, but a really thorough examination of their appropriateness would require computation far beyond the author's resources.‡

SYSTEMATIC SELECTION OF SAMPLING POINTS

3. Suppose that sample members are selected at points $t_1, t_2, \ldots, t_n$ in that order, and let $X_i = X(t_i)$ be the observed variables corrected for trend. Also let x_i denote the corresponding observations corrected for trend. Then the expected value of $(x_i - X_i)$ will be zero.

Let

$$E(x_i - X_i)^2 = \mu.$$

Then, expressed mathematically, the problem with which we shall be concerned here is that of estimating the random variable, I, where

$$I = (T-T_0)\bar{X} = \int_{T_0}^{T} X(t)\,dt,$$

from the random variables, $x_1, x_2, \ldots, x_n$.

* [Since $Ex(t) = 0$ the expected value of $\Sigma a_j x(t_j)$ is zero, not $\bar{X}$. The author means that the average of $\Sigma a_j x(t_j)$ *over all* t is equal to $\bar{X}$ which gives his condition that $\Sigma a_j = 1$.—ED.]

† If, in addition, we make the assumption (which is irrelevant to the subsequent development of this paper) that $X(t)$ is Gaussian, then $X(t)$ must be a stationary Gaussian-Markoff random function. It has been shown that such functions are everywhere continuous.

‡ [It was the Editor's intention to suggest that Dr Jones should include some of this experimental verification in his paper. The author assumes that $X(t)$ has the same mean, variance and correlation function for all t. His problem is analogous to the estimation of the stationary stochastic component of a continuous time-series defined at all points of time as the sum of three components, a linear trend, a random variable which possesses a correlogram decaying exponentially to zero and a 'superposed' error.—ED.]

Now since by 2 (A) and 2 (C)

$$E[(x_i - X_i)(x_j - X_j)] = 0 \quad (i \neq j), \tag{3·1}$$

$$E[X(s+t)X(s)] = \lambda e^{-q|t|}, \tag{3·2}$$

the correlation matrix of the vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is given by

$$\left.\begin{aligned} E(x_i x_j) &= \lambda e^{-q|t_i - t_j|} \quad (i \neq j) \\ &= \lambda + \mu \quad (i = j). \end{aligned}\right\} \tag{3·3}$$

If $\qquad P_i = e^{-qt_i}, \quad P_0 = e^{-qT_0}, \quad P_{n+1} = e^{-qT},$

then

$$E(\mathbf{xx}') = \lambda R + \mu I_n, \tag{3·4}$$

where

$$R = \begin{pmatrix} 1 & P_2/P_1 & P_3/P_1 & P_n/P_1 \\ P_2/P_1 & 1 & P_3/P_2 & P_n/P_2 \\ P_3/P_1 & P_3/P_2 & 1 & P_n/P_3 \\ \cdots & \cdots & \cdots & \cdots \\ P_n/P_1 & P_n/P_2 & \cdots & 1 \end{pmatrix}$$

and I_n is the unit matrix.*

Also, denoting $E(x_i I)$ by n_i, we have that

$$\begin{aligned} n_i &= E\left\{\int_{T_0}^{T} X(t)\, x_i\, dt\right\} \\ &= \int_{T_0}^{T} E[X(t)\, x_i]\, dt \\ &= \frac{\lambda}{q}\left(2 - \frac{P_i}{P_0} - \frac{P_{n+1}}{P_i}\right). \end{aligned} \tag{3·5}$$

THEOREM 1. Subject to the restriction that the sum of the weights, v_i, is $(T - T_0)$, the best estimate of I that can be obtained from any n samples is $\sum_{i=1}^{n} v_i x_i$, the sample values x_i being selected at points t_i $(t_i > t_{i-1})$, where

(a) $$v_1 = v_2 = \ldots = v_n = \frac{T - T_0}{n}. \tag{3·6}$$

(b) $$\begin{aligned} t_2 - t_1 &= t_3 - t_2 = \ldots = t_n - t_{n-1} \\ &= \frac{1}{q} \log_e \left\{ \left(k + \frac{\lambda - \mu}{2\lambda} \frac{T - T_0}{n} q\right) \Big/ \left(k - \frac{\lambda - \mu}{2\lambda} \frac{T - T_0}{n} q\right) \right\}. \end{aligned}$$

(c) $$t_1 - T_0 = T - t_n = -\frac{1}{q} \log_e \left(k - \frac{\lambda + \mu}{2\lambda} \frac{T - T_0}{n} q\right), \tag{3·7}$$

and k satisfies the equation

$$\left(k - \frac{\lambda + \mu}{2\lambda} \frac{T - T_0}{n} q\right)^{n+1} \Big/ \left(k + \frac{\lambda - \mu}{2\lambda} \frac{T - T_0}{n} q\right)^{n-1} = e^{-q(T - T_0)}. \tag{3·8}$$

Proof. The mean-square error of $\mathbf{v}'\mathbf{x}$ as an estimate of I will be

$$\begin{aligned} S &= E(I - \mathbf{v}'.\mathbf{x})^2 \\ &= \mathbf{v}'(\lambda R + \mu I_n)\mathbf{v} - 2\mathbf{v}'.\mathbf{n} + E(I^2). \end{aligned} \tag{3·9}$$

* The matrix notation used in this paper will be that of *The Theory of Canonical Matrices* by Turnbull and Aitken.

In order that $\mathbf{v}'\mathbf{x}$ be the best estimate, S must be a minimum for t_i and all possible v_i. Considering first the conditions that S should be a minimum by differentiating (3·9) partially with respect to t_i, we obtain by (3·3) and (3·5)

$$e^{-\alpha(t_i-T_0)}-e^{-\alpha(T-t_i)} = -\sum_{j=1}^{i-1} q\,e^{-\alpha(t_i-t_j)}\,v_j+\sum_{j=i+1}^{n} q\,e^{-\alpha(t_j-t_i)}\,v_j, \tag{3·10}$$

i.e.
$$\frac{P_i}{P_0}-\frac{P_{n+1}}{P_i} = -q\left\{\sum_{j=1}^{i-1} v_j\frac{P_i}{P_j}+\sum_{j=i+1}^{n} v_j\frac{P_j}{P_i}\right\} \quad (i = 1,\dots,n).$$

These n equations can be combined in the matrix form

$$\frac{P_i}{P_0}-\frac{P_{n+1}}{P_i} = [q(Q-Q')\,\mathbf{v}]_i \quad (i = 1,\dots,n), \tag{3·11}$$

where
$$Q = \begin{pmatrix} \frac{\lambda+\mu}{2\lambda} & \frac{P_2}{P_1} & \frac{P_3}{P_1} & \dots & \frac{P_n}{P_1} \\ 0 & \frac{\lambda+\mu}{2\lambda} & \frac{P_3}{P_2} & \dots & \frac{P_n}{P_2} \\ 0 & 0 & \frac{\lambda+\mu}{2\lambda} & \dots & \frac{P_n}{P_3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \dots & \frac{\lambda+\mu}{2\lambda} \end{pmatrix}.$$

Considering now possible variations of $\mathbf{v}$, the optimum values must satisfy

$$(\lambda R+\mu I_n)\,\mathbf{v}+k'.\mathbf{i} = \mathbf{n}, \tag{3·12}$$

i.e.
$$2k-\frac{P_{n+1}}{P_i}-\frac{P_i}{P_0} = [q(Q+Q')\,\mathbf{v}]_i \quad (i = 1,\dots,n), \tag{3·12'}$$

where $\mathbf{i}' = (1, 1, \dots, 1)$ and k', k are scalar constants.

The optimum estimate of I which can be made from n samples must thus satisfy (3·11) and (3·12'). Hence

$$k-\frac{P_{n+1}}{P_i} = [qQ\mathbf{v}]_i, \tag{3·13}$$

$$k-\frac{P_i}{P_0} = [qQ'\mathbf{v}]_i \quad (i = 1,\dots,n). \tag{3·14}$$

From (3·13) $$k(P_i-P_{i+1}) = q\left(\frac{\lambda+\mu}{2\lambda}P_i v_i+\frac{\lambda-\mu}{2\lambda}P_{i+1}v_{i+1}\right) \quad (i = 1,\dots,\overline{n-1}). \tag{3·15}$$

Similarly, from (3·14) $$k(P_i-P_{i+1}) = q\left(\frac{\lambda-\mu}{2\lambda}P_{i+1}v_i+\frac{\lambda+\mu}{2\lambda}P_i v_{i+1}\right). \tag{3·16}$$

Combining (3·15) and (3·16)

$$(v_i-v_{i+1})\left(P_{i+1}\frac{\lambda-\mu}{2\lambda}-P_i\frac{\lambda+\mu}{2\lambda}\right) = 0,$$

i.e.
$$v_i = v_{i+1} \quad (i = 1, 2,\dots,\overline{n-1}). \tag{3·6'}$$

Writing $v_1 = v_2 = \dots = v_n = (T-T_0)/n$ in the equations, we have

$$\frac{P_1}{P_2} = \frac{P_2}{P_3} = \dots = \frac{P_{n-1}}{P_n} = \left(k+\frac{\lambda-\mu}{2\lambda}\frac{T-T_0}{n}q\right)\Big/\left(k-\frac{\lambda+\mu}{2\lambda}\frac{T-T_0}{n}q\right). \tag{3·17}$$

Also we have
$$\frac{P_0}{P_1} = \frac{P_n}{P_{n+1}} = \left(k - \frac{\lambda+\mu}{2\lambda}\frac{T-T_0}{n}q\right)^{-1}. \tag{3·18}$$

Equations (3·17) and (3·18) lead immediately to results (3·7).

Finally, since
$$\frac{P_0}{P_{n+1}} = \frac{P_0}{P_1}\frac{P_1}{P_2}\cdots\frac{P_n}{P_{n+1}} = e^{q(T-T_0)},$$

k must be given by
$$\left(k + \frac{\lambda-\mu}{2\lambda}\frac{T-T_0}{n}q\right)^{n-1} \Big/ \left(k - \frac{\lambda+\mu}{2\lambda}\frac{T-T_0}{n}q\right)^{n+1} = e^{q(T-T_0)}. \tag{3·8}$$

Corollary. It may sometimes happen because the trend is known theoretically, or through some other circumstance, that it is unnecessary to impose the restriction
$$v_1 + v_2 + \ldots + v_n = T - T_0.$$

In this case we obtain by putting $k = 0$, $k = 1$ in the equations (3·13) and (3·14) that the optimum estimate of I is given by
$$v_1 = v_2 = \ldots = v_n = v \text{ (say)},$$
$$\frac{P_1}{P_2} = \frac{P_2}{P_3} = \ldots = \frac{P_{n+1}}{P_n} = 1 + \frac{\lambda-\mu}{2\lambda}vq \Big/ 1 - \frac{\lambda+\mu}{2\lambda}vq,$$
$$\frac{P_0}{P_1} = \frac{P_n}{P_{n+1}} = \left(1 - \frac{\lambda+\mu}{2\lambda}vq\right)^{-1},$$
$$\left(1 + \frac{\lambda-\mu}{2\lambda}vq\right)^{n-1} \Big/ \left(1 - \frac{\lambda+\mu}{2\lambda}vq\right)^{n+1} = e^{q(T-T_0)}.$$

Distribution of sample values—practical application

4. Two simple conclusions emerge from equations (3·6)–(3·8). They are that the optimum estimate of the 'average value' for a given sample size is one in which the same sample values are combined with equal weight $1/n$ and in which the distances between successive sample values are all equal, i.e. $t_2 - t_1 = t_3 - t_2 \ldots$. The distance of the first sample value from the end, $t_1 - T_0$, is governed by a rather more complicated equation, but it will be shown that $t_1 - T_0$ must lie between $\frac{1}{2}(t_2 - t_1)$ and $(t_2 - t_1)$. It is worth while to note, in passing, that though the formulae for $t_1, t_2, \ldots, t_n$ have been worked out on the assumption that the trend (§ 2) is simply a constant, *the formulae will still be accurate if the trend is a linear function of t.** In fact no other distribution of sample values is likely to decrease appreciably the error in the estimate of mean arising from the possible curvature of the trend.

It is only necessary to know the ratio $(t_1 - T_0)/(t_2 - t_1)$ to be able to specify the positions $t_1, t_2, \ldots, t_n$ completely.

THEOREM 2. If the distribution of sample values which gives the best estimate of the mean is at points $t_1, t_2, \ldots, t_n$ $(t_i > t_{i-1})$, then
$$t_2 - t_1 = t_{i+1} - t_i \quad (i = 2, 3, \ldots, n-1), \qquad t_1 - T_0 = T - t_n$$

* Suppose the trend could be expressed functionally as $A + Bt$. Then the truth of the above statement can readily be seen by considering the random function $X(t) = Z(t) - A - Bt$, and applying the formula,
$$E\left[\sum_{i=1}^{n} a_i z_i - \int_{T_0}^{T} Z(t)\,dt\right]^2 = E\left[\sum_{i=1}^{n} a_i x_i - \int_{T_0}^{T} X(t)\,dt\right]^2 + \left[\sum_{i=1}^{n} a_i t_i - \int_{T_0}^{T} t\,dt\right]^2.$$

and (*a*) $(t_1 - T_0)/(t_2 - t_1)$ is independent of λ and μ,

(*b*) $\tfrac{1}{2}(t_2 - t_1) \leqslant t_1 - T_0 \leqslant t_2 - t_1$, (4·1)

(*c*) $(t_1 - T_0)/(t_2 - t_1)$ can be determined very closely from the correlation of actual values, distance $(T - T_0)/n$ apart, i.e. from

$$e^{-q(T-T_0)/n}, \tag{4·2}$$

and is otherwise approximately independent of n.

Proof of (*a*). If $(T - T_0)\, q/n = r \qquad k - \tfrac{1}{2}\mu r/\lambda = K,$ (4·3)

$$p = e^{-q(t_i - t_{i-1})} \quad (i = 2, \ldots, n), \qquad p_0 = e^{-q(t_1 - T_0)}. \tag{4·4}$$

From (3·15)–(3·18)
$$p_0 = K - \tfrac{1}{2}r, \tag{4·5}$$
$$p = (K - \tfrac{1}{2}r)/(K + \tfrac{1}{2}r), \tag{4·6}$$
$$p_0^2 p^{n-1} = (K - \tfrac{1}{2}r)^{n+1}/(K + \tfrac{1}{2}r)^{n-1} = e^{-nr}. \tag{4·7}$$

The value of K is given by (4·7) and $(t_1 - T_0)/(t_2 - t_1)$ can be determined from (4·5) and (4·6). They are clearly independent of λ and μ.

Proof of (*b*). We have, from (4·6),

$$K - \tfrac{1}{2}r = rp/(1-p) = p_0. \tag{4·5'}$$

Hence by (4·7)
$$p^{n+1}/(1-p)^2 = e^{-nr}/r^2. \tag{4·8}$$

Now
$$p^{-\frac{1}{2}} - p^{\frac{1}{2}} \geqslant 2 \log_e 1/\sqrt{p} = \log_e 1/\sqrt{p}.$$

Hence
$$\frac{p}{(1-p)^2} \leqslant \frac{1}{(\log_e 1/p)^2}$$
$$\frac{e^{-nr}}{r^2} = \frac{p^{n+1}}{(1-p)^2} \leqslant \frac{e^{-n \log_e 1/p}}{(\log_e 1/p)^2}.$$

Now e^{-nr}/r^2 is a monotone decreasing function of r $(r > 0)$. Hence $\log_e 1/p \leqslant r$, and since $p_0^2 p^{n-1} = e^{-nr}$, it follows that $\log_e 1/p_0 \geqslant \tfrac{1}{2}r$. Hence

$$\frac{t_1 - T_0}{t_2 - t_1} \geqslant \tfrac{1}{2}. \tag{4·1'}$$

Also, since $1 - e^{-rn/(n+1)} \leqslant r$, by (4·7),

$$p(1-p^2)^{-2/(n+1)} = r^{-2/(n+1)}\, e^{-rn/(n+1)} \leqslant \frac{e^{-rn/(n+1)}}{(1 - e^{-rn/(n+1)})^{2/(n+1)}}.$$

Now $p(1-p^2)^{-2/(n+1)}$ is a monotone increasing function of p $(0 < p < 1)$. Hence $p \leqslant e^{-rn/(n+1)}$. But from (4·7) $p_0 \geqslant e^{-rn/(n+1)}$. So by (4·4)

$$\frac{t_1 - T_0}{t_2 - t_1} \leqslant 1. \tag{4·1''}$$

Thus (4·1′) and (4·1″) establish (4·1).

Proof of (*c*). To prove this result, we show that for a given r, $(t_1 - T_0)/(t_2 - t_1)$ is a monotone function of n. For

$$\frac{t_1 - T_0}{t_2 - t_1} = \frac{-\log (K - \tfrac{1}{2}r)}{-\log (K - \tfrac{1}{2}r) + \log (K + \tfrac{1}{2}r)}. \tag{4·9}$$

It follows from (4·2) that $K+\frac{1}{8}r>1$. Also K is a monotone function of n. Hence, when n increases $(t_1-T_0)/(t_2-t_1)$ decreases.

Some values of this ratio have been worked out and are presented in Table 1.* For $n=3$ the value is nearly the same as for n large and thus the ratio is nearly independent of n. This completes the proof of theorem 2.

Table 1 (*see text*)

$T-T_0$ = length of the interval sampled; n = number of sample members. L is the distance such that the correlation of true values, distance L apart, will be 0·5, i.e. is given by $e^{-\alpha L}=0{\cdot}5$.

$\dfrac{T-T_0}{nL}$	p = correlation between successive sample members	$\dfrac{t_1-T_0}{t_2-t_1}$ (n large)	$\dfrac{t_1-T_0}{t_2-t_1}$ ($n=3$)
0·5	0·707	0·51	0·51
1·0	0·500	0·53	0·51
2·0	0·250	0·56	0·54
3·0	0·125	0·58	0·56
4·0	0·062	0·61	0·58
5·0	0·031	0·63	0·60
7·0	0·008	0·67	0·65
10·0	0·001	0·72	0·69
15·0	$0{\cdot}0^43$	0·77	0·74
20·0	$0{\cdot}0^51$	0·82	0·78

We may expand the ratio $(t_1-T_0)/(t_2-t_1)$ in terms of r. It will be sufficient to consider the expansion when n is large as, by theorem 2, the result is nearly independent of n. We have

$$\frac{t_1-T_0}{t_2-t_1}=\frac{\log\{pr/(1-p)\}}{\log p}=\frac{r+\log\{r/(1-p)\}}{r}$$

$$=\frac{1}{2}+\frac{r}{24}+O(r^3). \tag{4·10}$$

A similar expansion in terms of the parameter $1-e^{-r}$, which is nearly equal to $1-p$, is

$$\frac{1}{2}+\frac{1-e^{-r}}{24}+\frac{(1-e^{-r})^2}{48}+O(1-e^{-r})^3. \tag{4·11}$$

We cannot expect to have anything but a rough idea of r before an experiment is conducted. Fortunately, this is all that will be required. It will be seen from Table 1 that if $e^{-r}>0{\cdot}25$ ($r>2{\cdot}8$), no accuracy will be lost if the sample members are placed at distances $(T-T_0)/2n$, $3(T-T_0)/2n$, etc., from T_0. In other words, if the correlation between successive sample members is likely to be appreciable, we divide the whole interval into n equal sections, and place one member at each of the mid-points of the sections. Again, if the number of sample members is large, then the loss of accuracy resulting from placing the samples at distances $(T-T_0)/2n$, $3(T-T_0)/2n$ from T_0 should be small.

* [*Editorial note.* Dr Jones appears to have derived the last column of Table 1, first by determining r by successive approximation from p according to the relation $e^{-nr}/r^2=p^{n+1}/(1-p)^2$, and then by the use of

$$(t_1-T_0)/(t_2-t_1)=\log\{pr/(1-p)\}/\log p.$$

The approximation of (4·10) is not very useful for small p because r is then large. It may be shown that for large n the ratio $(t_1-T_0)/(t_2-t_1)$ tends to $1-\{\log(1-p)-\log|\log p|\}/\log p$ which obviates the necessity for determining r.]

CONTINUATION OF DR JONES'S PAPER

BY M. G. KENDALL

1. In his original version of the foregoing paper Dr Jones included a Part II on the estimation of mean-square error. His treatment was open to misunderstanding and his methods of estimating constants unnecessarily elaborate. The most unfortunate accident of his death prevented a discussion of these matters with him. The editors felt that his work was important enough to justify publication and the foregoing paper appears accordingly. In this continuation I have re-examined the work in his Second Part and have added some comments of my own. It was, the Editors felt, scarcely within their province to substitute this work for that of Dr Jones, and to publish it under his name, although some of the methods are due to him.

2. In the first place I wish to comment on Dr Jones's theorem 2 and its consequences. He has shown in his theorem 1 that the optimum distribution of sampling points is obtained when they are equidistant, and essentially his second theorem is concerned with the relative length of the two end-segments of the range of t as divided by the sampling points.

Write
$$\alpha = \frac{t_1 - T_0}{t_2 - t_1}. \tag{1}$$

Then since $p = e^{-q(t_2-t_1)}$ and $p_0 = e^{-q(t_1-T_0)}$, $\alpha = \log p_0/\log p$. From Dr Jones's equation (4·7) we have $p_0^2 p^{n-1} = e^{-nr}$ and thus it follows that
$$\alpha = \tfrac{1}{2} - \frac{n}{2}\left(1 + \frac{r}{\log p}\right). \tag{2}$$

Theorem 2 shows that
$$\tfrac{1}{2} \leqslant \alpha \leqslant 1. \tag{3}$$

Thus, as n becomes larger $1 + r/\log p$ must tend to zero, or p tends to e^{-r}. Furthermore, since $r = (T - T_0)q/n$, r must tend to zero (and p accordingly to unity) as n tends to infinity, *provided that* $(T - T_0)q$ *remains constant*. Since q is an (unknown) positive constant of the system, this is equivalent to the proviso that $T - T_0$ must remain constant.

3. I think Dr Jones's Table 1 tends to obscure this point. If we keep the interval $T - T_0$ constant, then as n increases p tends to unity, as we should expect from considerations of continuity. *Per contra*, for large n the interval increases as p becomes smaller. The increase in n for fixed p then does not correspond to a denser distribution of sample points but to the extension of the sampled strip. In such a case it seems an unnecessary refinement to discuss at great length the end intervals when the total interval is tending to infinity. Dr Jones may have been thinking here, not of his example of a strip of soil, but of meteorological variation which goes on indefinitely.

4. Let $(T - T_0)q = d$. Since p tends to unity with increasing n for fixed d we may write
$$1 - p = \frac{a_1}{n} + \frac{a_2}{n^2} + \frac{a_3}{n^3} + \frac{a_4}{n^4} + \dots. \tag{4}$$

On substituting in
$$e^{-nr}(1-p)^2 = r^2 p^{n+1} \tag{5}$$

and identifying coefficients we find
$$a_1 = d, \quad a_2 = -\tfrac{1}{2}d^2, \quad a_3 = \frac{(2d+3)}{d+2}\frac{d^3}{12}, \quad a_4 = -\frac{d^5}{24(d+2)}. \tag{6}$$

Hence we find
$$1+\frac{r}{\log p} = \frac{a_2}{n^2}+\frac{1}{3}\frac{a_1^3}{dn^3}+O(n^{-3})$$
$$= -\frac{1}{6}\frac{d^2}{n^2}+O(n^{-3}). \tag{7}$$

Thus to order n^{-1}
$$\alpha = \frac{1}{2}+\frac{1}{12}\frac{d^2}{n}, \tag{8}$$

and so, as n tends to infinity α tends to $\frac{1}{2}$ in all cases where the interval d is fixed. It appears to me, therefore, that the general rule for large samples is to divide the interval into n equal parts and place one sample member at the middle of each. The discussion for small p is only appropriate when the sample is large and the interval is also large, in which case the location of the first sample member can scarcely be of much practical interest.

5. If we keep p fixed and let n tend to infinity, suppose that
$$1+\frac{r}{\log p} = \frac{b_1}{n}+\frac{b_2}{n^2}+\dots. \tag{9}$$

Such an expansion is possible because, from (2), $1+r/\log p$ must tend to zero with increasing n in all cases. On substitution in (5) we find to order n^{-1}
$$\frac{p^{1+b_1}(\log p)^2}{(1-p)^2}\left(1-\frac{2b_1}{n}\right) = 1-\frac{b_2}{n}\log p,$$

whence
$$b_1 = -1+\frac{2}{\log p}\{\log(1-p)-\log|\log p|\}, \tag{10}$$
$$b_2 = 2b_1/\log p. \tag{11}$$

Thus
$$\alpha = 1-\frac{1}{\log p}\{\log(1-p)-\log|\log p|\}. \tag{12}$$

It happens to be true that as p tends to unity this value of α tends to $\frac{1}{2}$.

6. It may be noted in passing that expansion by these methods (d fixed, n large) leads to the series
$$\alpha = \frac{1}{2}+\frac{r}{24}\frac{d}{d+2}+O(r^3), \tag{13}$$

which is not the same as Dr Jones's equation (4·10) unless we let d become large.

7. If the trend is constant or linear the mean-square error of our estimate for $\bar{Z}$ is the same as that for $\bar{X}$. We then require
$$E\left(\frac{1}{n}\Sigma x-\bar{X}\right)^2 = \frac{1}{n^2}E(\Sigma x)^2-\frac{2}{n}E(\Sigma x\bar{X})+E(\bar{X})^2. \tag{14}$$

Now
$$E(\bar{X}^2) = \frac{1}{(T-T_0)^2}\int_{T_0}^{T}\int_{T_0}^{T} E\{X(t)\,X(u)\}\,dt\,du$$
$$= \frac{1}{(T-T_0)^2}\int_{T_0}^{T}\int_{T_0}^{T} \lambda e^{-q|t-u|}\,dt\,du$$
$$= \frac{2\lambda}{(T-T_0)q}-\frac{2\lambda(1-e^{-q(T-T_0)})}{(T-T_0)^2q^2}. \tag{15}$$

Using Dr Jones's (3·5) we have

$$\frac{2}{n}E(\Sigma x\bar{X}) = \frac{2}{n}\sum_j \frac{\lambda}{q}(2-p_0p^{j-1}-p_0p^{n-j})$$

$$= \frac{2\lambda}{nq}\left(2n - \frac{2p_0(1-p^n)}{1-p}\right)$$

$$= \frac{4\lambda}{(T-T_0)q}\left\{1-\frac{p(1-p^n)q(T-T_0)}{n^2(1-p)^2}\right\}, \tag{16}$$

$$E\left(\frac{1}{n}\Sigma x\right)^2 = \frac{1}{n^2}(\Sigma x^2+\Sigma x_i x_j) \quad (i \neq j)$$

$$= \frac{\mu}{n}+\lambda\{1+p+p^2+\ldots+p^{n-1}+p+1+p+\ldots+p^{n-2}+\ldots+p^{n-1}+p^{n-2}+\ldots+1\}$$

$$= \frac{\mu}{n}+\frac{\lambda(1+p)}{n(1-p)}-\frac{2\lambda p(1-p^n)}{n^2(1-p)^2}. \tag{17}$$

Hence, on substitution in (9), we find

$$S = E\left(\frac{1}{n}\Sigma x-\bar{X}\right)^2 = \frac{\mu}{n}+\lambda\left\{\frac{1+p}{n(1-p)}-\frac{2}{q(T-T_0)}\right\}$$

$$+2\lambda\left\{\frac{p(1-p^n)}{n^2(1-p)^2}-\frac{1-e^{-q(T-T_0)}}{(T-T_0)^2q^2}\right\}. \tag{18}$$

Since $$\frac{p^{n+1}}{n^2(1-p)^2} = \frac{e^{-nr}}{n^2r^2} = \frac{e^{-q(T-T_0)}}{(T-T_0)q^2},$$

this simplifies slightly to

$$S = \frac{\mu}{n}+\lambda\left\{\frac{1+p}{n(1-p)}-\frac{2}{q(T-T_0)}\right\}+\frac{2\lambda}{n^2}\left\{\frac{p}{(1-p)^2}-\frac{n^2}{q^2(T-T_0)^2}\right\}. \tag{19}$$

8. Dr Jones, having reached this expression by a slightly different route, proceeded to argue that the last term was of order n^{-2} and could be neglected, whereas the second term was of order n^{-1}. Noting that $q(T-T_0)$ was of order $n\log 1/p$ he was led to the expression, for large n,

$$S = \frac{1}{n}\{\mu+\lambda F(p)\}, \tag{20}$$

where $$F(p) = \frac{1+p}{1-p}+\frac{2}{\log p}. \tag{21}$$

He gave the attached Table 2 for $F(p)$ and emphasized that even for low values of p such as 0·01 the mean-square error (apart from the term in μ) was only half of the value for $p = 0$. He inferred that even a trace of correlation between successive sample members would seriously affect the mean-square error.

9. Now if we hold p constant and let n tend to infinity, formulae (20) and (21) result. But in doing so, as I have already pointed out, we are extending the interval to infinity, not concentrating the sample more densely. In these circumstances it is not so surprising that small

correlations should affect the mean-square error substantially. It appears to me to be more relevant to the problem to consider the limiting form of (19) when n tends to infinity but $d = q(T-T_0)$ remains fixed. In such a case we get quite different results.

10. Consider the first term in braces on the right-hand side of (19).
Substituting from (4) we find

$$\frac{1+p}{n(1-p)}-\frac{2}{d}=\left(2-\frac{a_1}{n}-\frac{a_2}{n^2}-\frac{a_3}{n^3}-\dots\right)\Big/\left(a_1+\frac{a_2}{n}+\frac{a_3}{n^2}+\dots\right)-\frac{2}{d},$$

which, to order n^{-2}, reduces to
$$\frac{(d+3)\,d}{6(d+2)\,n^2}, \tag{22}$$

which is of order n^{-2}, not n^{-1}.

Similarly for the last expression in (19)

$$\frac{p}{n^2(1-p)^2}-\frac{1}{d^2}=-\frac{d}{12(d+2)\,n^2}. \tag{23}$$

To order n^{-1} we then have merely
$$S=\frac{\mu}{n}, \tag{24}$$

or, to order n^{-2},
$$S=\frac{\mu}{n}+\frac{\lambda d}{6n^2}. \tag{25}$$

Table 2 (*see text*)

p	$F(p)$	p	$F(p)$	p	$F(p)$
0·0	1·000	0·06	0·417	0·35	0·172
$0{\cdot}0^51$	0·855	0·07	0·398	0·40	0·151
0·001	0·712	0·08	0·382	0·45	0·137
0·005	0·633	0·09	0·367	0·50	0·115
0·01	0·586	0·10	0·354	0·60	0·085
0·02	0·530	0·15	0·299	0·70	0·059
0·03	0·491	0·20	0·257	0·80	0·037
0·04	0·462	0·25	0·224	0·90	0·018
0·05	0·438	0·30	0·196	1·00	0·000

11. The appearance of the term in μ is to be expected; it represents the variance of superposed error of observation. But the result that the remaining part of the mean-square error is of order n^{-2} is at first sight surprising. One is so accustomed in statistical work to a sampling variance of order n^{-1} that anything of lower order requires some explanation.

It is here that we must remember that our problem is not the determination of the mean-square error or the sampling variance of n independent observations. On the contrary, under our assumptions, the correlation between neighbouring sample members tends to unity with increasing n. The variation of such members among themselves is of order n^{-2} as may be verified from (17), neglecting the term in μ. Since we choose our function $\bar{X}$ so as to fit these observations as closely as possible it is not, after all, surprising that the average difference square of $\frac{1}{n}\Sigma(x)$ and $\bar{X}$ is also of order n^{-2}.

12. There is one essential discontinuity in the situation, however, which is worth noticing. If $q(T-T_0)$ remains finite then if n tends to infinity, p must tend to unity and the series (apart from error of observation) is continuous. However large q may be this is true. But if q is itself infinite the series is not necessarily continuous, the observations are independent and we revert to the usual situation of n independent observations. Thus for any q, however large, (25) is correct; but for q infinite it is not. Looking back to (19) we see that (since p in this case is zero)

$$S = \frac{\mu}{n} + \frac{\lambda}{n},$$

which is what we should expect.

13. From (19) we see that (without approximation) the mean-square error S depends on μ, λ and p; on the known quantities $T-T_0$ and n; and on q, which is known when p is known and the interval between successive observations is known. In practice we may require to estimate μ, λ and p. In the limiting case which Dr Jones considered, leading to (20) we also require μ, λ and p. In the limiting case which I am considering, leading to (25), we require only μ and λ.

Dr Jones proceeded by dismissing μ and considering the estimation of λ and p from the expectations of powers of differences $\Sigma(x_j - x_{j-1})^2$. This is really equivalent to using the serial correlations of the observations.

We have, since $E(x) = 0$,

$$E(x^2) = \mu + \lambda, \quad E(x_j x_{j+1}) = \lambda p, \quad E(x_j x_{j+2}) = \lambda p^2, \text{ etc.}$$

If we take the observed serial covariances of the observations (freed from trend) as estimators of the corresponding expectations we then have, if the serial correlations are r_1, r_2, etc.

$$p = \frac{r_2}{r_1}, \tag{26}$$

$$\lambda = \frac{r_1^2}{r_2} \operatorname{var} x, \tag{27}$$

$$\mu = \operatorname{var} x \left(1 - \frac{r_1^2}{r_2}\right). \tag{28}$$

These seem to me to be the simplest equations of estimation which one is likely to find. If we may assume that $\mu = 0$ we have the more reliable forms

$$p = r_1, \tag{29}$$

$$\lambda = \operatorname{var} x. \tag{30}$$

These equations are the usual ones for estimating constants in a Markoff series.

14. One final comment. The relative simplicity of Dr Jones's result that the optimum distribution of sample points is equidistantly along the interval is a little deceptive. It is natural to consider the more general problem when the autocorrelations along the series are not necessarily decaying according to an exponential law, but have any form permissible for a continuous random process. An intuitive approach might suggest that the weights and distances of the observations should be equal because there is no obvious reason to the contrary; but any conclusion of this kind would be quite wrong. If we denote the autocorrelation

function of the series by $\rho(t)$ and the sample points are $t_1 \ldots t_n$ in that order the general equations corresponding to Dr Jones's (3·11) and (3·12) are

$$\left.\begin{aligned}\left(1+\frac{\mu}{\lambda}\right)v_1+v_2\rho(t_2-t_1)+\ldots+v_n\rho(t_n-t_1) &= E\{x(t_1)\bar{X}\},\\ v_1\rho(t_2-t_1)+v_2\left(1+\frac{\mu}{\lambda}\right)+\ldots+v_n\rho(t_n-t_2) &= E\{x(t_2)\bar{X}\},\\ \\ v_1\rho(t_n-t_1)+v_2\rho(t_n-t_2)+\ldots+v_n\left(1+\frac{\mu}{\lambda}\right) &= E\{x(t_n)\bar{X}\},\end{aligned}\right\} \tag{31}$$

$$\left.\begin{aligned}0+v_2\rho'(t_2-t_1)+\ldots+v_n\rho'(t_n-t_1) &= E'\{x(t_1)\bar{X}\},\\ -v_1\rho'(t_2-t_1)+0+\ldots+v_n\rho'(t_n-t_2) &= E'\{x(t_2)\bar{X}\},\\ \\ -v_1\rho'(t_n-t_1)+v_2\rho'(t_n-t_2)+\ldots+0 &= E'\{x(t_n)\bar{X}\},\end{aligned}\right\} \tag{32}$$

where the primes denote differentiation. I cannot see any general tractable solution to these equations. Consideration of some particular cases suggests that general solutions would be rather involved. For instance, if the autocorrelation function is a sine curve (corresponding to sinusoidal periodicity in the original series) a set of equidistant observations might be the worst possible if the distance between them was equal to the period of the system. Again, if the autocorrelation function decays to zero in distance l and the observations are so sparse as to be farther than l apart they are independent and their position is indeterminate within limits. Further research on this subject is needed.

THE PRECISION OF OBSERVED VALUES OF SMALL FREQUENCIES

BY J. B. S. HALDANE, F.R.S.

In recent genetical work numerous observers have recorded the frequencies of rare events, notably mutations. It has been realized that it is misleading to state the observed frequencies with their standard errors, since the distribution is decidedly skew. Various devices have been suggested to avoid this difficulty. But so far as I know it has not been pointed out that, when the frequency is small, its cube root is almost normally distributed. This will be proved and applied to actual observations.

Let a rare event be observed in a out of n trials, where n is much greater than a^2. Let x be the true value of the frequency, whose observed value is $p = a/n$. Let the *a priori* distribution of x be

$$dF = \phi(x)\,dx.$$

Let the probability distribution, after the observation has been made, be

$$dF = f(x)\,dx,$$

and let $x = y^3$.

Then for given values of n and x, the probability of a is

$$\binom{n}{a} x^a (1-x)^{n-a}.$$

Hence for given values of n and a, the distribution of x is

$$dF = f(x)\,dx = \frac{x^a(1-x)^{n-a}\phi(x)\,dx}{\int_0^1 x^a(1-x)^{n-a}\phi(x)\,dx}.$$

If we assume that all values of x are equiprobable, $\phi(x) = 1$, and

$$\bar{x} = \frac{(n+1)!}{a!\,(n-a)!}\int_0^1 x^{a+1}(1-x)^{n-a}\,dx = \frac{a+1}{n+2}.$$

This value should of course be a/n. As I have previously remarked (Haldane, 1932) and as Jeffreys (1948) has shown in greater detail, the assumption that $\phi(x) = 1$ introduces a bias. It is also contrary to common sense. If we are trying to estimate a mutation rate, we know *a priori* that it will almost certainly be less than 10^{-3} and greater than 10^{-20}. In a particular case we might perhaps guess that such a rate would be about as likely to lie between 10^{-5} and 10^{-6} as between 10^{-6} and 10^{-7}. In other words, when x is small it is more nearly true that all values of $\log x$ are equiprobable than that all values of x are equiprobable. This would imply that $\phi(x) = c/x$ in the region considered. However, this cannot continue to be true when x is sufficiently small. If we wished to state a plausible general form for the *a priori* distribution of x it might be somewhat as follows:

$$F = k \quad (x = 0),$$

$$dF = \frac{C}{(x+\epsilon)(1+\epsilon-x)} \quad (0 < x < 1),$$

$$F = k \quad (x = 1),$$

where k is some number less than $\frac{1}{2}$ expressing the possibility that x may prove to be zero or unity,

$$C = \frac{(1-2k)(1+2\epsilon)}{2\log(1+\epsilon^{-1})},$$

and ϵ is a very small number, perhaps of the order of 10^{-100}, expressing the fact that exceedingly rare events are relatively infrequent. If the universe is finite in space and in time, and if there is a minimum time in which an event can occur, it might imply that there is no sense in discussing events which have no appreciable probability of ever occurring.

For practical purposes, however, so long as we know that a exceeds zero, and is less than n, that is to say, that the event considered is possible and so is its converse, we can take $\phi(x) = \dfrac{C}{x(1-x)}$ without appreciable error. We then have

$$dF = \frac{(n-1)!}{(a-1)!\,(n-a-1)!} x^{a-1}(1-x)^{n-a-1}\,dx. \tag{1}$$

This is a Pearsonian Type I distribution, and

$$\overline{x^r} = \frac{(n-1)!\,(a+r-1)!}{(n+r-1)!\,(a-1)!}.$$

Thus $\bar{x} = a/n$, as it should be, $\overline{x^2} = \dfrac{a(a+1)}{n(n+1)}$, etc. When $an^{-\frac{1}{2}}$ is small, this approximates very closely to the Type III distribution

$$dF = \frac{e^{-nx}x^{a-1}}{(a-1)!}dx. \tag{2}$$

Now Wilson & Hilferty (1931) showed that the cube root of χ^2 is almost normally distributed; and the same transformation will almost normalize many Type III distributions.

The standard form of this type, referred to its mode, is

$$dF = C\left(1+\frac{x}{a}\right)^{\gamma a} e^{-\gamma x}\,dx.$$

It is more convenient to change the origin to the point where the probability becomes zero, and write

$$dF = \frac{\gamma^c x^{c-1}\,dx}{\Gamma(c)\,e^{\gamma x}},$$

where $$c = 1+p = 1+\gamma a = 4/\beta_1.$$

$\kappa_r = (r-1)!\,c\gamma^{-r}$, so the mean is $c\gamma^{-1}$ and the moments about it are

$\mu_2 = c\gamma^{-2}$, $\mu_4 = (3c^2+6c)\gamma^{-4}$, $\mu_6 = (15c^3+130c^2+120c)\gamma^{-6}$, $\mu_8 = (105c^4+\dots)\gamma^{-8}$,
$\mu_3 = 2c\gamma^{-3}$, $\mu_5 = (20c^2+24c)\gamma^{-5}$, $\mu_7 = (210c^3+924c^2+720c)\gamma^{-7}$, $\mu_9 = (2520c^4+\dots)\gamma^{-9}$.

Let $x = c\gamma^{-1}+z$, so that $\overline{z^r} = \mu_r$ and $y = (\gamma x/c)^{\frac{1}{3}}$. Then

$$y = \left(1+\frac{\gamma z}{c}\right)^{\frac{1}{3}}$$

and
$$\begin{aligned}\overline{y^r} &= 1+\tfrac{1}{3}r\frac{\gamma\bar{z}}{c}+\binom{\frac{1}{3}r}{2}\frac{\gamma^2\overline{z^2}}{c^2}+\binom{\frac{1}{3}r}{3}\frac{\gamma^3\overline{z^3}}{c^3}+\dots\\ &= 1+\frac{r(r-3)}{18c}+\frac{r(r-1)(r-3)(r-6)}{2(18c)^2}+\frac{r^2(r-3)^2(r-6)(r-9)}{6(18c)^3}\\ &\quad+\frac{r(r-3)(r-6)(r-9)(r-12)(5r^3-30r^2+15r+18)}{120(18c)^4}+O(c^{-5}).\end{aligned}$$

Or, putting $t = \dfrac{1}{9c}$,

$$\begin{aligned}
\bar{y} &= 1-t+\tfrac{10}{3}t^3+\tfrac{11}{3}t^4+O(t^5),\\
\overline{y^2} &= 1-t+t^2+\tfrac{7}{3}t^3-\tfrac{28}{3}t^4+O(t^5),\\
\overline{y^3} &= 1,\\
\overline{y^4} &= 1+2t-3t^2+\tfrac{10}{3}t^3+\tfrac{41}{3}t^4+O(t^5),\\
\overline{y^5} &= 1+5t-5t^2+\tfrac{25}{3}t^3+\tfrac{14}{3}t^4+O(t^5),\\
\overline{y^6} &= 1+9t.
\end{aligned}$$

Hence the cumulants of the distribution of y are

$$\left.\begin{aligned}
\kappa_1 &= 1-t+\tfrac{10}{3}t^3+\tfrac{11}{3}t^4+O(t^5),\\
\kappa_2 &= t-\tfrac{13}{3}t^3-10t^4+O(t^5),\\
\kappa_3 &= 4t^3+16t^4+O(t^5),\\
\kappa_4 &= -2t^3-16t^4+O(t^5),\\
\kappa_5 &= 8t^4+O(t^5),\\
\kappa_6 &= -55t^4+O(t^5),
\end{aligned}\right\} \tag{3}$$

or

$$\begin{aligned}
\sigma &= \frac{1}{(9c)^{\frac{1}{2}}}\left[1-\frac{13}{6(9c)^2}-\frac{5}{(9c)^3}+O(c^{-4})\right],\\
\gamma_1 &= \frac{4}{(9c)^{\frac{3}{2}}}\left[1+\frac{4}{9c}+O(c^{-2})\right],\\
\gamma_2 &= \frac{-2}{9c}\left[1+\frac{8}{9c}+O(c^{-2})\right],\\
\gamma_3 &= \frac{8}{(9c)^{\frac{3}{2}}}+O(c^{-\frac{5}{2}}),\\
\gamma_4 &= \frac{-55}{9c}+O(c^{-2}).
\end{aligned}$$

Thus provided $9c$, or $9(1+p)$, is large, the approximation to normality, up to the sixth moment, is satisfactory. But it is of no value when p is negative, that is to say, the curve is J-shaped. (Here p is of course the parameter used in specifying Type III distributions, and not the observed frequency value.)

To apply these formulae to the distribution of y, we have only to put $a = c$, and to multiply κ_r by $p^{\frac{1}{3}r}$. We thus find

$$\left.\begin{aligned}
\kappa_1 &= p^{\frac{1}{3}}[1-\tfrac{1}{9}a^{-1}+\tfrac{10}{2187}a^{-3}+O(a^{-4})], & \kappa_3 &= p[\tfrac{4}{729}a^{-3}+O(a^{-4})],\\
\kappa_2 &= p^{\frac{2}{3}}[\tfrac{1}{9}a^{-1}-\tfrac{13}{2187}a^{-3}+O(a^{-4})], & \kappa_4 &= p^{\frac{4}{3}}[\tfrac{2}{729}a^{-3}+O(a^{-4})], \quad \text{etc.}
\end{aligned}\right\} \tag{4}$$

Thus $\sigma = p^{\frac{1}{3}}/3a^{\frac{1}{2}}$, $\gamma_1 = \frac{4}{27}a^{-\frac{3}{2}}$, $\gamma_2 = \frac{2}{9}a^{-1}$, all approximately.

The terms involving a^{-3} in the mean and standard error may be safely neglected in practice. Even when $a = 1$, the former is only 0·013 of the standard error. If we take (1) as our distribution of x, a term of order n^{-1} must be added to those of order a^{-3}. This also can be safely neglected.

Thus we find that y is almost normally distributed with mean $(1-\frac{1}{9}a^{-1})\,p^{\frac{1}{3}}$, and standard deviation $p^{\frac{1}{3}}/3a^{\frac{1}{2}}$. For example, if $n = 1000$, $a = 8$, $p = 0{\cdot}008$, x is by no means normally distributed about 0·008, for $\beta_1 = 0{\cdot}5$ and $\beta_2 = 3{\cdot}75$. But y is very nearly normally distributed

about $0{\cdot}2\times\frac{71}{72}$ or $0{\cdot}1972$ with standard deviation $\frac{1}{120}$ or $0{\cdot}0083$, with $\beta_1 = 0{\cdot}00004$, and $\beta_2 = 3{\cdot}028$. The method of Haldane (1938) would give an even better fit if $a > 10$.

Two examples will be given showing how the method can be actually used.

Muller (1928, p. 311) found 13 lethal genes in 1034 X-chromosomes of flies kept at 27° C., and 5 in 840 X-chromosomes of flies kept at 19·5° C. Thus corrected values of y_1 and y_2 are:—

$$y_1' = \left(\frac{13}{1034}\right)^{\frac{1}{3}}\left(\frac{1}{1-9\times 13}\right) = 0{\cdot}23054 \pm 0{\cdot}02150, \quad y_2' = 0{\cdot}17720 \pm 0{\cdot}02702.$$

$y_1' - y_2' = 0{\cdot}05334$, which is 1·55 times its standard error of 0·03452. The difference is therefore rather more significant than Muller, who used the usual formula, believed.

Again Muller (1940) obtained 7 translocations in 3366 flies with a dose of 375 r., and 56 in 2223 flies with a dose of 1500 r. The question at issue was as follows: 'the frequency may be proportional to the dosage, to its $\frac{3}{2}$th power or to its square. With which, if any, of these hypotheses are the observed results consistent?'

$$y_1' = 0{\cdot}125616 \pm 0{\cdot}016081, \quad y_2' = 0{\cdot}29256 \pm 0{\cdot}01306.$$

We therefore compare y_2' with

$$2^{\frac{2}{3}}y_1' = 0{\cdot}19940 \pm 0{\cdot}02559, \quad 2y_1' = 0{\cdot}25123 \pm 0{\cdot}03216, \quad 2^{\frac{4}{3}}y_1' = 0{\cdot}31653 \pm 0{\cdot}04052.$$

The differences are respectively 3·25, 1·19 and 0·56 times their standard errors, so either of the latter two hypotheses is admissible.

It is perhaps worth remarking that, if the emendation of the classical inverse probability distribution be rejected, and the calculation made according to Bayes's hypothesis, the cube root of the frequency is still almost normally distributed. It is also true that if the frequency of a rare event is estimated by the method described by Haldane (1945) when the observations cease when a fixed number m of rare events have occurred, the estimated frequency being $(m-1)/(n-1)$, where n is the total number of observations, the cube root of the estimate is almost normally distributed. Here too the cube root may be used with advantage in comparing different estimates.

I have to thank Prof. E. S. Pearson for valuable criticism.

Summary

When an event is rare, the distribution of the cube root of the frequency round the cube root of the estimate is much more nearly normal than the distribution of the true frequency round the estimate.

REFERENCES

Haldane, J. B. S. (1932). A note on inverse probability. *Proc. Camb. Phil. Soc.* **28**, 55–61.

Haldane, J. B. S. (1938). The approximate normalization of a class of frequency distributions. *Biometrika*, **29**, 392–404.

Haldane, J. B. S. (1945). On a method of estimating frequencies. *Biometrika*, **33**, 222–5.

Jeffreys, H. (1948). *Theory of Probability*. Oxford University Press.

Muller, H. J. (1928). The measurement of gene mutation rate in *Drosophila*, its high variability and its dependence on temperature. *Genetics*, **13**, 274–357.

Muller, H. J. (1940). An analysis of the process of structural change in chromosomes of *Drosophila*. *J. Genet.* **40**, 1–66.

Wilson, E. B. & Hilferty, M. M. (1931). The distribution of Chi-square. *Proc. Nat. Acad. Sci., Wash.*, **17**, 684–88.

NOTE ON PROFESSOR HALDANE'S PAPER REGARDING THE TREATMENT OF RARE EVENTS

By E. S. PEARSON

In the preceding paper Prof. Haldane has suggested a method of handling certain problems involving the occurrence of rare events, by introducing the cube-root transformation. His method of attack involves the use of the concept of inverse probability, so that his final, closely normal distribution is the posterior probability distribution of y, the cube root of the unknown probability x of the occurrence of the event. From this point of view x, and therefore y, are continuous variables. If we attack the problem without an appeal to inverse probability, we are concerned with the probability distribution of a, the observed sample frequency, for a given x. Since a is a discontinuous random variable which, in the problems considered, may only assume the first few integer values $0, 1, 2, \ldots$, the cube-root transformation would here clearly introduce some awkward problems of discontinuity. It seems of interest to consider how the examples which Haldane gives could be dealt with by the direct method without the need of substituting approximate standard errors or, indeed, of assuming that a very skew distribution is normal.

In both Haldane's examples we are concerned with a comparison of the results of two experiments. In a first sample of size n_1 an event occurs on a_1 occasions, the chance of occurrence being x_1; similarly for the second sample we have a_2, n_2 and x_2. We then ask whether the results are consistent with the hypothesis that $x_1 = \kappa x_2$, where, for the first example $\kappa = 1$ and for the second has to be taken successively as $\frac{1}{4}$, $(\frac{1}{4})^{\frac{3}{2}}$ and $(\frac{1}{4})^2$.

The assumption which I shall make is that in the problems considered x and $1/n$ are sufficiently small to justify the use of the Poisson series in place of the binomial expansion $(\overline{1-x+x})^n$. If this is the case and we write

$$m_i = n_i x_i \quad (i = 1, 2),$$

the chance of the observed event may be written

$$p(a_1, a_2 \mid m_1, m_2) = e^{-(m_1+m_2)}\frac{m_1^{a_1} m_2^{a_2}}{a_1!\,a_2!} = \frac{e^{-\mu}\mu^{a_1+a_2}}{(a_1+a_2)!} \times \frac{(a_1+a_2)!}{a_1!\,a_2!}\lambda^{a_1}(1-\lambda)^{a_2}, \tag{1}$$

where

$$\mu = m_1 + m_2, \quad \lambda = \frac{m_1}{m_1+m_2}. \tag{2}$$

For a fixed value of $r = a_1 + a_2$, the relative distribution of a_1 follows a binomial, $(\overline{1-\lambda+\lambda})^r$, where λ is specified by hypothesis. Thus if the hypothesis is that $m_1 = m_2$, then $\lambda = \frac{1}{2}$, but in general

$$\lambda = \frac{n_1 x_1}{n_1 x_1 + n_2 x_2} = \frac{n_1 \kappa}{n_1 \kappa + n_2}, \tag{3}$$

where κ is the ratio of the true or hypothetical chances, x_1/x_2.

As in the case of the more general problem of the analysis of 2×2 tables, where the conditional distribution is a hypergeometric series, there may be some difference of opinion on the way in which the result is used.

(1) We may consider that the whole answer lies in the conditional distribution of a_1 (or a_2) for fixed r, in the sense that we need only ask whether the partition of the observed events

into a_1 occurring in the first sample and a_2 in the second is consistent with our hypothesis as to κ. The answer is obtained in terms of probability by summing the tail terms of the binomial $(\overline{1-\lambda}+\lambda)^r$. This approach appears a natural one to take when two treatments have been randomly assigned among n_1+n_2 individuals.

(2) We may wish to set the observed result against the two-dimensioned distribution of a_1 and a_2 obtainable in unrestricted sampling, i.e. without the condition that $r = a_1+a_2$ is fixed. This distribution depends on the value of $\mu = m_1+m_2$ which is not specified by the hypothesis, but it is possible to obtain upper limits to significance levels; this problem was considered by Przyborowski & Wilenski (1939) for the case $\lambda = \frac{1}{2}$, and further tables for some other special cases were circulated during the war within the Ministry of Supply (Barnard, 1944; Allinson, 1944).

(3) Still avoiding the condition that r is fixed, we may note that for unrestricted sampling

$$u = \frac{a_1 - r\lambda}{\sqrt{\{r\lambda(1-\lambda)\}}} \tag{4}$$

is a random variable which, on the hypothesis tested, has a zero expectation and unit variance. Its sampling distribution is composite, being the sum of a number of binomials combined with weights depending on the unknown value of $\mu = m_1+m_2$. However, if λ is not too different from $\frac{1}{2}$, this distribution will not be very far from the normal even when dealing with small frequencies. A calculation of the ratio u will therefore often provide the broad answer needed in practice.

These points are illustrated on Haldane's examples.

Example 1. $a_1 = 13$, $n_1 = 1034$; $a_2 = 5$, $n_2 = 840$; $r = 18$. In this problem, the hypothesis tested is that $x_1 = x_2$, so that $\kappa = 1$ and $\lambda = n_1/(n_1+n_2) = 0{\cdot}5518$. On the assumption that the flies were randomly divided into the two temperature groups, the null hypothesis is that 18 out of the 1874 would have produced progeny of a type which showed they carried a lethal gene at whichever temperature they were kept. Using the binomial approximation to the hypergeometric, a_1 would assume values of $0, 1, 2, \ldots, 18$ with probabilities given by the expansion of $(0{\cdot}4482+0{\cdot}5518)^{18}$. The chance that $a_1 \geqslant 13$ is given by the Incomplete Beta Function Ratio

$$I_\lambda(a_1, r-a_1+1) = I_{0{\cdot}5518}(13, 6) = 0{\cdot}1107.$$

An approximation to this chance can be obtained from the integral under the normal curve having

$$\text{Mean} = r\lambda = 9{\cdot}932, \quad \text{S.D.} = \sqrt{\{r\lambda(1-\lambda)\}} = 2{\cdot}110,$$

using the correction for continuity. We then get the ratio $(12{\cdot}5-9{\cdot}932)/2{\cdot}110 = 1{\cdot}217$, corresponding to a chance of $0{\cdot}1118$.

If we take the approach of (3) above, avoiding restriction to the conditional set $r = 18$, then a correction for continuity is not appropriate and we find

$$u = (13-9{\cdot}932)/2{\cdot}110 = 1{\cdot}45,$$

a ratio to be compared with Haldane's $1{\cdot}55$, obtained from the posterior distribution of $x_1^{\frac{1}{2}} - x_2^{\frac{1}{2}}$.

Example 2. Here $a_1 = 7$, $n_1 = 3366$; $a_2 = 56$, $n_2 = 2223$; $r = 63$. Three hypotheses are examined, (*a*), (*b*) and (*c*), and as two of these involve the assumption that the chance x of translocation varies with dose, the randomization approach to the conditional distribution is less clear.* I think I should here base my conclusions on the values of the ratio u. Relevant

* In this case, with $a_2 = 56$, we are beyond the range of the Tables of the Incomplete Beta Function.

figures are given below, the means and standard deviations being for the binomials $(\overline{1-\lambda}+\lambda)^{63}$. It will be seen that the values of $(a_1-r\lambda)/\sqrt{\{r\lambda(1-\lambda)\}}$ are not very different from Haldane's ratios, and that the same conclusions would be reached using either method of approach.

Hypothesis	Mean $a_1 \mid r$	S.D. of $a_1 \mid r$	$u=(a_1\text{-mean})/\text{S.D.}$	Haldane's ratio
(a) $\kappa=\frac{1}{4}$, $\lambda=0{\cdot}2746$	17·300	3·542	−2·91	−3·24
(b) $\kappa=\frac{1}{8}$, $\lambda=0{\cdot}1591$	10·026	2·904	−1·04	−1·19
(c) $\kappa=\frac{1}{16}$, $\lambda=0{\cdot}0865$	5·447	2·231	0·70	0·56

To sum up:

(1) if conditions justify us in using the Poisson series to represent the binomial distribution of a, given x, which will generally be the case with 'rare' events,

(2) if we are content to base our answer on the conditional distribution of a_1 for fixed $r = a_1 + a_2$,

the direct method of attack in this two-sample problem provides a test involving a binomial distribution. The exact answer in terms of a significance level requires the calculation of the sum of tail terms of the binomial, which can be obtained directly from the Tables of the Incomplete Beta Function if both a_1 and $a_2+1 \leqslant 50$. Alternatively, if we do not wish to restrict variation to the conditional set, we may obtain a less precise answer by referring $u = (a_1-r\lambda)/\sqrt{\{r\lambda(1-\lambda)\}}$ to the normal integral, a procedure which will be satisfactory if λ is not too far from 0·5.

Haldane's solution with its continuous normal distribution is obviously attractive, but it has involved the introduction of the theory of inverse probability.

REFERENCES

ALLINSON, V. A. (1944). Technical Report Q.C./R/21 (Ministry of Supply).
BARNARD, G. A. (1944). Technical Report Q.C./R/18 (Ministry of Supply).
HALDANE, J. B. S. (1948). *Biometrika*, **35**, 297.
PRZYBOROWSKI, J. & WILENSKI, H. (1939). *Biometrika*, **31**, 313.

A FURTHER NOTE ON THE MEAN DEVIATION

By H. J. GODWIN, *University College of Swansea*

1. *Introduction.* In a previous paper (1945), I obtained the distribution of the estimate of mean deviation obtained from samples from a normal population; the method of derivation was an algebraic transformation of the sample space. In the present note I explain the geometrical significance of the result, and obtain a method of computing the moments of the distribution. Finally, I discuss various approximations to the distribution.

2. *Geometrical prologue.* The fundamental geometrical entity which will occur here is the regular simplex in k dimensions—a figure whose vertices are $(k+1)$ points each equidistant from the remainder. (Particular cases are the equilateral triangle and the regular tetrahedron.) These points lie on a hypersphere whose centre is the centroid of the simplex. Any k of the points form a regular simplex of $(k-1)$ dimensions; the join of the centroid of this to the centroid of the whole passes through the remaining vertex and is perpendicular to the space of the $(k-1)$-dimensional simplex. Since the centroid divides the perpendicular from a vertex to the opposite face in the ratio $k:1$, the angle made with each other by the (equally inclined) lines from the centroid to the vertices is $\arccos(-1/k)$. If the length of side of the simplex is a, then each vertex is at a distance $a\left(\frac{k}{2(k+1)}\right)^{\frac{1}{2}}$ from the centroid and, by a suitable choice of axes through the centroid, the vertices are

$$-\frac{a}{\sqrt{\{2k(k+1)\}}},\quad -\frac{a}{\sqrt{\{2(k-1)k\}}},\quad \ldots,\quad -\frac{a}{\sqrt{\{2(k-r+1)(k-r+2)\}}},$$
$$a\sqrt{\left(\frac{k-r}{2(k-r+1)}\right)},\quad 0,\ 0,\ \ldots,\ 0 \quad (r=0,1,\ldots,k). \tag{1}$$

The vertices lie in sets of k on the $(k+1)$ bounding hyperplanes

$$-\sqrt{\left(\frac{2}{k(k+1)}\right)}x_1-\sqrt{\left(\frac{2}{k(k-1)}\right)}x_2-\ldots-\sqrt{\left(\frac{2}{(k+1-r)(k+2-r)}\right)}x_r$$
$$+\sqrt{\left(\frac{2(k-r)}{k+1-r}\right)}x_{r+1}+\frac{a}{k+1}=0 \quad (r=0,1,\ldots,k). \tag{2}$$

The defining relations for the interior of the simplex are that the left-hand sides of (2) should be not less than zero.

We now find the value of the integral of $\exp\{-\frac{1}{2}(x_1^2+x_2^2+\ldots+x_k^2)\}$ taken through the interior of the simplex. Let this be $E(k,a)$. By joining the centroid of the simplex to the edges of any one of the $(k-1)$-dimensional simplexes in the bounding hyperplanes we obtain $(k+1)$ equal regions and the integral of $\exp\{-\frac{1}{2}(x_1^2+x_2^2+\ldots+x_k^2)\}$ through such a region is $E(k,a)/(k+1)$. A section of a region by a hyperplane parallel to the base hyperplane and at a perpendicular distance x from the centroid is a simplex of side $x\sqrt{\{2k(k+1)\}}$. Hence

$$\frac{E(k,a)}{k+1}=\int_0^{a/\sqrt{\{2k(k+1)\}}}E[k-1,x\sqrt{\{2k(k+1)\}}]\,e^{-\frac{1}{2}x^2}\,dx.$$

It follows, by induction, that $\quad E(k,a)=\sqrt{(k+1)}\,G_k(a/\sqrt{2}),\qquad(3)$

where the function G_k is as defined in my 1945 paper. [This is, perhaps, the appropriate place in which to remark that this function is related to a function defined by McKay (1935) in dealing with the distribution of deviations from the greatest observation in a sample; in fact,

his $F_n(x) = \sqrt{n}\, G_{n-1}(nx)/(2\pi)^{\frac{1}{2}(n-1)}$. I am indebted to Dr H. O. Hartley for bringing this to my notice.]

From (3) it follows that $G_k(\infty) = \dfrac{1}{\sqrt{(k+1)}} E(k,\infty) = \dfrac{(2\pi)^{\frac{1}{2}k}}{\sqrt{(k+1)}}$.

3. *Distribution of the mean deviation.* Let a sample of n from the normal population (supposed, without loss of generality, to have zero mean and unit variance) be $x_1, x_2, \ldots, x_n$. Let the mean of the sample be $\bar{x}$, and let $d_i = x_i - \bar{x}$ ($i = 1, 2, \ldots, n$). Consider the case in which k of the d's are negative and the rest positive. Then $1 \leqslant k \leqslant n-1$. By suitably numbering the d's the negative ones may be taken as $d_1, \ldots, d_k$; the mean deviation of the sample is then

$$m = n^{-1}[-d_1 - d_2 - \ldots - d_k + d_{k+1} + \ldots + d_n].$$

We now transform the sample space of the x's by means of the matrix equation $Y = TX$, where X, Y are column vectors $(x_1, \ldots, x_n)$, $(y_1, \ldots, y_n)$ and T is the orthogonal matrix:

$$\begin{array}{c} \overbrace{\hphantom{xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx}}^{k \text{ columns}} \\ \left[\begin{array}{cccc}
\frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\
-\frac{n-k}{\sqrt{\{kn(n-k)\}}} & -\frac{n-k}{\sqrt{\{kn(n-k)\}}} & \cdots & -\frac{n-k}{\sqrt{\{kn(n-k)\}}} \\
-\frac{k-1}{\sqrt{\{k(k-1)\}}} & \frac{1}{\sqrt{\{k(k-1)\}}} & \cdots & \frac{1}{\sqrt{\{k(k-1)\}}} \\
0 & -\frac{k-2}{\sqrt{\{(k-1)(k-2)\}}} & \cdots & \frac{1}{\sqrt{\{(k-1)(k-2)\}}} \\
\cdots & \cdots & \cdots & \cdots \\
0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\
0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots \\
0 & 0 & \cdots & 0
\end{array}\right. \end{array}
\begin{array}{l} \\ \\ \\ \left.\begin{array}{c} \\ \\ \\ \\ \\ \end{array}\right\} k-1 \text{ rows} \\ \left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} n-k-1 \text{ rows} \end{array}$$

$$\begin{array}{c} \overbrace{\hphantom{xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx}}^{n-k \text{ columns}} \\ \left.\begin{array}{cccc}
\frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\
\frac{k}{\sqrt{\{kn(n-k)\}}} & \frac{k}{\sqrt{\{kn(n-k)\}}} & \cdots & \frac{k}{\sqrt{\{kn(n-k)\}}} \\
0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots \\
0 & 0 & \cdots & 0 \\
\frac{n-k-1}{\sqrt{\{(n-k)(n-k-1)\}}} & -\frac{1}{\sqrt{\{(n-k)(n-k-1)\}}} & \cdots & -\frac{1}{\sqrt{\{(n-k)(n-k-1)\}}} \\
0 & \frac{n-k-2}{\sqrt{\{(n-k-1)(n-k-2)\}}} & \cdots & -\frac{1}{\sqrt{\{(n-k-1)(n-k-2)\}}} \\
\cdots & \cdots & \cdots & \cdots \\
0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}
\end{array}\right] \end{array}
\begin{array}{l} \\ \\ \\ \left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} k-1 \text{ rows} \\ \left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} n-k-1 \text{ rows} \end{array}$$

Then $y_1 = \sqrt{n}\,\bar{x}$, and $y_2 = \frac{1}{2}n^{\frac{1}{2}}m/\sqrt{\{k(n-k)\}}$. Since $X = T'Y$, T' being the transposed matrix of T, the d's can easily be expressed in terms of the y's, and this gives, using equations (2), that the appropriate region in the space of the co-ordinates $y_3, \ldots, y_{k+1}$ is a regular simplex of side $y_2\{2k(n-k)/n\}^{\frac{1}{2}} = \frac{1}{2}nm$, and in the space of the co-ordinates $y_{k+2}, \ldots, y_n$, a regular simplex also of side $\frac{1}{2}nm$. The frequency function of m is $f(m)\,dm$, which is the integral of $2\pi^{-\frac{1}{2}n}\exp\{-\frac{1}{2}(x_1^2+\ldots+x_n^2)\}$, taken through the region in which the mean deviation lies between m and $m+dm$. Since T is an orthogonal matrix, $(x_1^2+\ldots+x_n^2) = (y_1^2+\ldots+y_n^2)$, and we may integrate over the variable y_1 (which is independent of m) from $-\infty$ to ∞. The contribution to $f(m)\,dm$ from the region described above is

$$\frac{1}{(2\pi)^{\frac{1}{2}(n-1)}}e^{-\frac{1}{2}y_2^2}\,dy_2\sqrt{\{k(n-k)\}}\,G_{k-1}(\tfrac{1}{2}nm)\,G_{n-k-1}(\tfrac{1}{2}nm)$$

$$= \frac{n^{\frac{1}{2}}}{2(2\pi)^{\frac{1}{2}(n-1)}}\exp\left[-\frac{n^3m^2}{8k(n-k)}\right]G_{k-1}(\tfrac{1}{2}nm)\,G_{n-k-1}(\tfrac{1}{2}nm)\,dm.$$

Now, for a given value of k, there are ${}^nC_k$ ways of allocating the k negative d's and k may vary from 1 to $n-1$. Hence

$$f(m)\,dm = \frac{n^{\frac{1}{2}}}{2(2\pi)^{\frac{1}{2}(n-1)}}\sum_{k=1}^{n-1}{}^nC_k\exp-\left[\frac{n^3m^2}{8k(n-k)}\right]G_{k-1}(\tfrac{1}{2}nm)\,G_{n-k-1}(\tfrac{1}{2}nm)\,dm,$$

as obtained before (1945).

Exact descriptions of the sample space for small values of n may be of interest. For $n = 2$, m is constant on two straight lines both perpendicular to the lines $\bar{x}$ = constant. For $n = 3$, m is constant on a cylinder, whose axis is perpendicular to the planes $\bar{x}$ = constant, and whose cross-section is a regular hexagon (i.e. ${}^3C_1+{}^3C_2$ equal lines). For $n = 4$, m is constant, for any space $\bar{x}$ = constant, on the surface of a cube of side $4m$, whose corners have been 'filed off' to give equilateral triangles of side $2\sqrt{2}\,m$. The surface thus consists of 6 (i.e. ${}^4C_2$) squares and 8 (i.e. ${}^4C_1+{}^4C_3$) equilateral triangles.

4. *Moments of the distribution.* We now derive a recurrence relation from which the moments of the distribution can be calculated. We denote

$$\int_0^\infty\left(\sum_{k=r}^{n-r}{}^{n-2r}C_{k-r}\,t^s\exp\left[-\frac{nt^2}{2k(n-k)}\right]G_{k-1}(t)\,G_{n-k-1}(t)\right)dt$$

by $I(n,r,s)$ $(s \geqslant 0)$. (The summation is from 1 to $n-1$ when $r = 0$.) For $s \geqslant 1$ we have, on integrating by parts, suitably grouping terms and, in one case, changing the summation variable from k to $k+1$, the successive equations

$$I(n,r,s) = \int_0^\infty\sum{}^{n-2r}C_{k-r}\frac{k(n-k)}{n}\exp\left[-\frac{nt^2}{2k(n-k)}\right]$$

$$\times\left\{(s-1)\,t^{s-2}G_{k-1}(t)\,G_{n-k-1}(t)+t^{s-1}\exp\left[-\frac{t^2}{2k(k-1)}\right]G_{k-2}(t)\,G_{n-k-1}(t)\right.$$

$$\left.+\,t^{s-1}\exp\left[-\frac{t^2}{2(n-k-1)(n-k)}\right]G_{k-1}(t)\,G_{n-k-2}(t)\right\}dt$$

$$= \int_0^\infty\sum{}^{n-2r}C_{k-r}\frac{r(n-r)+(k-r)(n-k-r)}{n}(s-1)\,t^{s-2}\exp\left[-\frac{nt^2}{2k(n-k)}\right]G_{k-1}(t)\,G_{n-k-1}(t)\,dt$$

$$+\int_0^\infty\sum\exp\left[-\frac{(n-1)\,t^2}{2k(n-k-1)}\right]G_{k-1}(t)\,G_{n-k-2}(t)\,t^{s-1}$$

$$\times\left[\frac{k(n-k)}{n}{}^{n-2r}C_{k-r}+\frac{(k+1)(n-k-1)}{n}{}^{n-2r}C_{k-r+1}\right]dt,$$

and, since
$$\frac{k(n-k)}{n}{}^{n-2r}C_{k-r}+\frac{(k+1)(n-k-1)}{n}{}^{n-2r}C_{k-r+1}$$
$$=\frac{(n-2r)(n-2r-1)}{n}{}^{n-2r-1}C_{k-r}+\frac{r(n-r)}{n}{}^{n-2r+1}C_{k-r+1},$$
we have
$$I(n,r,s)=\frac{r(n-r)(s-1)}{n}I(n,r,s-2)+\frac{(n-2r)(n-2r-1)(s-1)}{n}I(n,r+1,s-2)$$
$$+\frac{(n-2r)(n-2r-1)}{n}I(n-1,r,s-1)+\frac{r(n-r)}{n}I(n-1,r-1,s-1). \quad (4)$$

Now the sth moment of the distribution of m is $n^{\frac{1}{2}}(2\pi)^{-\frac{1}{2}(n-1)}(2/n)^s I(n,0,s)$ and by the aid of (4) we can express $I(n,0,s)$ in terms of certain of the $I(n,r,0)$, $I(n-1,r,0)$, etc. (For a given value of s, we need values of r for which $0\leqslant 2r\leqslant s$.) $I(n,r,0)$ is the integral from 0 to ∞ of certain terms in the frequency function of m, viz. ${}^{n-2r}C_{k-r}$ of those arising when k of the d's are negative ($r\leqslant k\leqslant n-r$). To get the whole frequency function ${}^{n}C_k$ terms are taken because that is the number of ways of choosing k negative d's. But if r of the d's are fixed as positive and r as negative (i.e. a certain part only of the sample space is considered), then the remaining $(k-r)$ negative d's can be chosen in ${}^{n-2r}C_{k-r}$ ways. Hence $I(n,r,0)$ is the integral of $(2\pi n)^{-\frac{1}{2}}\exp[-\frac{1}{2}\Sigma x^2]$ over this restricted region of the sample space. In particular, $I(n,0,0)=(2\pi)^{\frac{1}{2}(n-1)}n^{-\frac{1}{2}}$, a result which also follows from the frequency function of m.

To evaluate the integral it is convenient to transform the sample space by the matrix T, k being taken equal to 1. (This does not imply as much as before about the relation of the mean to the other observations, since we shall now only restrict the signs of two of the d's.) For $r=1$, this gives the definition of the restricted region to be
$$z_2>0, \quad \frac{z_2}{\sqrt{\{n(n-1)\}}}+z_3\sqrt{\left(\frac{n-2}{n-1}\right)}>0.$$
The value of the integral of $(2\pi n)^{-\frac{1}{2}}\exp[-\frac{1}{2}\Sigma z^2]$ over this region bears to its value over all space the ratio (i) $\frac{1}{2}\pi+\sin^{-1}[1/(n-1)]$ to (ii) 2π, since (i) is the angle between the two hyperplanes bounding the region. Hence
$$\frac{I(n,1,0)}{I(n,0,0)}=\frac{1}{4}+\frac{1}{2\pi}\sin^{-1}\frac{1}{n-1}.$$

The calculation of $I(n,2,0)$ involves a much more complicated integration. The result is that
$$I(n,2,0)=\frac{(2\pi)^{\frac{1}{2}(n-1)}}{\sqrt{n}}\left[\frac{1}{16}+\frac{\sin^{-1}[1/(n-1)]+\sin^{-1}[1/(n-3)]}{8\pi}\right.$$
$$+\frac{5\{(\sin^{-1}[1/(n-1)])(\sin^{-1}[1/(n-3)])\}}{4\pi^2}$$
$$\left.-\frac{1}{\pi^2}\int_0^{\sin^{-1}[1/(n-3)]}\tan^{-1}\left(\frac{2-(n-1)(n-4)\tan^2\phi}{n(n-3)}\right)^{\frac{1}{2}}d\phi\right].$$
The result can be put in several forms, but this is the one which I personally have found most useful for computation.

For larger r the $I(n,r,0)$ would presumably be even more complicated; the two given are sufficient to determine the first five moments of the distribution. Since the d's tend to

independence for large n and the assignment of sign to a d reduces the sample space by one-half, we have that

$$I(n,r,0) = \frac{(2\pi)^{\frac{1}{2}(n-1)}}{\sqrt{n}}\left(\frac{1}{4^r}+O\left(\frac{1}{n}\right)\right).$$

By the use of the recurrence relation (4) we can now find the moments and constants of shape of the distribution of m; these agree with the expressions obtained by Geary (1936) (see also some of his results quoted by Pearson (1945)), except that the coefficient of n^{-6} in the expansion of m_4'' (formula (22), 1936 paper) should be $(51-352a^2+(2136/5)a^4)$ and not $(51-352a^2+427a^4)$. This mistake causes the coefficient of n^{-6} in the expansion of $m_4''-m_4'$ (formula (24), same paper) to be 0·033578 instead of the correct 0·114635, and that of ν^{-6} in the expansion of λ_4 (formula (4), 1945) to be $-0{\cdot}120003$ instead of the correct $-0{\cdot}038946$.

We also have the new result

$$\mu_5 = \frac{0{\cdot}792218}{n^3}+\frac{0{\cdot}023893}{n^4}+\frac{0{\cdot}097967}{n^5}+\frac{0{\cdot}133384}{n^6}+\ldots.$$

5. *Approximations to the distribution.* The exact calculation of tables of the probability integral of m is lengthy, and the labour increases as n does. Consequently it seemed worth while to investigate the accuracy of approximations to the distribution which would involve less work in their computation. This investigation was of an empirical nature and consisted of comparing the values given for the percentage points of the distribution of m for $n = 10$ with the true values calculated by Hartley (1945). It was assumed that the approximation which is most accurate for that value of n will also be the most accurate for other values. The approximations considered were

(*a*) the normal distribution with the same first two moments as m;
(*b*) the distribution of the form $Km^{\alpha}e^{-\beta m^2}$ with the same first two moments as m;
(*c*) the Pearson curve (in fact, Type I) with the same first four moments as m;
(*d*) functions of m chosen so that their distributions are more nearly normal than that of m.

Of these methods, (*d*) is the most useful; however, a brief description of the others may be of interest.

(*a*), as may be seen by the comparison in Table 2 following my earlier paper, is useless for small n. Comparing it with (*d*) for $n = 1000$, it is found that the percentage points are given with an error of about 0·01 in the extreme values (0·1 and 99·9 %).

(*b*), when fitted to sample size 10 gives the 99 % point with error 0·01 and the 99·9 % point with error 0·02.

(*c*), gives percentage points for sample size 10 correct to three places of decimals for cumulative probabilities between 1 and 99·8 %. The computation is laborious, however, and gets heavier for large n, owing to the difficulty of calculating values of the incomplete B-function for large p, q.

(*d*) is an application of a method due to Haldane (1937). The distributions of the functions $y = (m/\overline{m})^k$ and $z = (1+(m-\overline{m})/g)^h$ are considered, where k, g and h are chosen so that for the distribution of y, β_1 is $O(n^{-3})$, while for the distribution of z, β_1 is $O(n^{-3})$ and β_2-3 is $O(n^{-2})$. To find k, g and h, powers of y and z are expanded in powers of $(m-\overline{m})$, and their products with $f(m)\,dm$ integrated with m ranging from 0 to ∞. The moments of y and z are thus found in terms of the moments of m and can be arranged as power series in n^{-1}. (In a review of Haldane's paper, Neyman (1938) points out that this method is not rigorous as the series in $(m-\overline{m})$ are divergent for some values of m; this does not, however, necessarily imply that the expansions of the moments of y and z in terms of n^{-1}, as far as is needed, are

false. An investigation of the remainder terms of these series would be of interest.) k, g and h are now found so as to make the moments of y and z of the required order. To find g and h a knowledge of the sixth moment of m is needed. To avoid calculation of $I(n, 3, 0)$ it was assumed that, for the distribution of m,

$$\mu_6/(15\mu_2^3) = 1 + O(n^{-1});$$

this seems reasonable in view of the fact that $\mu_4/(3\mu_2^2)$, $\mu_5/(10\mu_2\mu_3)$ are both $1 + O(n^{-1})$. The results are that

$$y = (m/\overline{m})^{0{\cdot}5609}$$

is approximately normally distributed with mean

$$1 - \frac{0{\cdot}0703}{n} \quad \frac{0{\cdot}0714}{n^2} \quad 0{\cdot}0818$$

and standard deviation $$\frac{0{\cdot}42375}{\sqrt{n}}\left(1 + \frac{0{\cdot}4808}{n} + \frac{0{\cdot}4099}{n^2} + \dots\right),$$

while $$z = \left(1 + \frac{m - \overline{m}}{0{\cdot}6005}\right)^{0{\cdot}6695}$$

is approximately normally distributed with mean

$$1 - \frac{0{\cdot}1115}{n} - \frac{0{\cdot}0239}{n^2} - \frac{0{\cdot}0135}{n^3} - \dots$$

and standard deviation $$\frac{0{\cdot}67205}{\sqrt{n}}\left[1 + \frac{0{\cdot}0219}{n} + \frac{0{\cdot}0082}{n^2} + \dots\right].$$

For sample size 10, y gives percentage points differing by amounts up to 0·006 (for 99 %) from the true values, while z gives values differing by not more than 0·001. The two methods give the same values (to three places of decimals) when n is greater than 35. The labour of calculation does not increase with n, and for a percentage point which is frequently needed, the equation giving y or z in terms of m can be inverted to give the desired value directly as a power series in $n^{-\frac{1}{2}}$.

REFERENCES

GEARY, R. C. (1936). *Biometrika*, **28**, 295–305.

GODWIN, H. J. (1945). *Biometrika*, **33**, 254–6.

HALDANE, J. B. S. (1937). *Biometrika*, **29**, 392–404.

HARTLEY, H. O. (1945). *Biometrika*, **33**, 257–65.

McKAY, A. T. (1935). *Biometrika*, **27**, 466–71.

NEYMAN, J. (1938). *Zbl. Math.* **18**, 257.

PEARSON, E. S. (1945). *Biometrika*, **33**, 252–3 (Editorial Note).

A NOTE ON THE ASYMPTOTIC DISTRIBUTION OF RANGE

BY D. R. COX, *Wool Industries Research Association*

1. INTRODUCTION AND SUMMARY

In a recent paper Elfving (1947) has given an asymptotic form for the distribution of range in large samples.

In the present note, two other methods of obtaining an asymptotic form for the range distribution are discussed, which have the advantage of being expressed directly in terms of the range, while Elfving's form involved a non-linear transformation of range.

Numerical results are given for a normal population comparing the exact distribution of range with the two approximations discussed here, and with Elfving's approximation.

2. DERIVATION FROM FISHER AND TIPPETT'S RESULTS ON THE DISTRIBUTION OF EXTREMES

Fisher & Tippett (1928) have obtained results for the distribution of the least and greatest members of large random samples. Now in large samples it is clear that least and greatest values are effectively independent, and so a form for the asymptotic distribution of range can be derived by integration from the joint distribution of least and greatest values.*

Consider random samples of size n taken from a population whose frequency function is $\phi(x)$ and whose distribution function is $\Phi(x)$. Define $x_n^{(1)}$ and $x_n^{(2)}$ by

$$\Phi(x_n^{(1)}) = 1-\frac{1}{n}, \tag{1}$$

$$\Phi(x_n^{(2)}) = \frac{1}{n}. \tag{2}$$

Let

$$y_1 = n(x_1-x_n^{(1)})\,\phi(x_n^{(1)}), \tag{3}$$

$$y_2 = n(x_2-x_n^{(2)})\,\phi(x_n^{(2)}), \tag{4}$$

where x_1 and x_2 are the greatest and least members of the sample.

Then Fisher & Tippett have shown that if $\phi(x)$ tends to zero exponentially or faster as x tends to infinity, the limiting frequency functions of y_1 and y_2 are

$$\exp(-y_1-e^{-y_1}) \quad \text{and} \quad \exp(y_2-e^{y_2}).$$

Thus the limiting joint-frequency function of y_1 and y_2 is

$$\exp(-y_1-e^{-y_1}+y_2-e^{y_2}),$$

and so if

$$W = y_1-y_2, \tag{5}$$

the limiting frequency function of W is

$$\int_{-\infty}^{\infty} \exp(-W-e^{-W-y_2}-e^{y_2})\,dy_2 = 2e^{-W}K_0(2e^{-\frac{1}{2}W}). \tag{6}$$

In equation (6), $K_0(x)$ is a modified Bessel function of the second kind (Watson, 1944, p. 78).

* Since I wrote this paper, it has been pointed out to me that this method has recently been discussed at length by E. J. Gumbel (1947). In the case of a symmetrical distribution, W of equation (5) below is termed by Gumbel the 'reduced range' and denoted by R. For the case of a normal population he compares the exact probability levels of the sample range with those derived from the asymptotic distribution of equation (6) for which he has calculated certain values of the probability integral. He has not, however, considered the rather closer approximation which I discuss in § 3 below.

Suppose now that the basic distribution is symmetrical and has mean zero, so that

$$x_n^{(1)} = -x_n^{(2)},$$

and
$$\phi(x_n^{(1)}) = \phi(x_n^{(2)}).$$

Then if w_n is the sample range

$$w_n = 2x_n^{(2)} + \frac{W}{n\phi(x_n^{(2)})}. \tag{7}$$

Thus (6) leads immediately to a form for the asymptotic frequency function of range.

It can be shown directly from (6), or inferred from Fisher & Tippett's work, that the mean and standard deviation of W in the distribution (6), are

$$\overline{W} = 2\gamma, \tag{8}$$

$$\sigma_W = \frac{\pi}{\sqrt{3}}, \tag{9}$$

where γ is Euler's constant.

Also
$$\beta_1 = 0\cdot806, \tag{10}$$

and
$$\beta_2 = 4\cdot2. \tag{11}$$

If we combine these results with equation (7), expressing the range w_n in terms of W, we find that for large n

$$\overline{w}_n = 2x_n^{(2)} + \frac{2\gamma}{n\phi(x_n^{(2)})}, \tag{12}$$

$$\sigma_{w_n} = \frac{\pi}{\sqrt{3}\, n\phi(x_n^{(2)})}. \tag{13}$$

The limiting β coefficients of w_n are the same as those of W.

A numerical comparison of these results with the exact results for the normal distribution is given in § 4.

3. Asymptotic form by the method of steepest descents

The exact expression for the frequency function of range involves an integral. For large n, this integral can be evaluated by the method of steepest descents and this leads to a further form for the asymptotic frequency function of range.

The frequency function, $f_n(w_n)$, of the range w_n in samples of size n, is known to be

$$\begin{aligned} f_n(w_n) &= n(n-1)\int_{-\infty}^{\infty} \phi(u-\tfrac{1}{2}w_n)\phi(u+\tfrac{1}{2}w_n)\,[\Phi(u+\tfrac{1}{2}w_n)-\Phi(u-\tfrac{1}{2}w_n)]^{n-2}\,du \\ &= n(n-1)\int_{-\infty}^{\infty} \phi(u-\tfrac{1}{2}w_n)\phi(u+\tfrac{1}{2}w_n)\exp\,[(n-2)\,\psi(u,w_n)]\,du, \end{aligned} \tag{14}$$

where
$$\psi(u,w_n) = \log\,[\Phi(u+\tfrac{1}{2}w_n)-\Phi(u-\tfrac{1}{2}w_n)]. \tag{15}$$

It will now be supposed for simplicity that $\phi(x)$ is a symmetrical unimodal frequency distribution with mean zero; that is, that

$$\phi(x) = \phi(-x), \qquad \phi'(x) \neq 0 \quad (x \neq 0).$$

Then for fixed w_n, the function of u, $\psi(u,w_n)$ has a single maximum at $u = 0$. The integral (14) is thus of the form that can be evaluated by the method of steepest descents (Watson, 1944, p. 235).

The basic idea is that as n tends to infinity, by far the greatest portion of the integral (14) comes from the neighbourhood of $u = 0$.

Write $$\psi(u, w_n) = \psi(0, w_n) - \tfrac{1}{2}t^2, \tag{16}$$
and then it is found that
$$f_n(w_n) = n(n-1)\exp[(n-2)\psi(0,w_n)]\int_{-\infty}^{\infty}\phi(u-\tfrac{1}{2}w_n)\phi(u+\tfrac{1}{2}w_n)\frac{du}{dt}\exp\left[-\frac{(n-2)t^2}{2}\right]dt. \tag{17}$$

The equation (16) can be solved for u as a power series in t, and so the expression
$$\phi(u-\tfrac{1}{2}w_n)\phi(u+\tfrac{1}{2}w_n)\frac{du}{dt}$$
can be obtained as a function of t.

It is found that
$$\phi(u-\tfrac{1}{2}w_n)\phi(u+\tfrac{1}{2}w_n)\frac{du}{dt} = \frac{[\phi(\frac{1}{2}w_n)]^2}{[-\psi''(0,w_n)]^{\frac{1}{2}}}[1 + A_2(w_n)t^2 + \dots].$$

In this expression $A_2(w_n)$ is a complicated function of $\Phi(\tfrac{1}{2}w_n)$, $\phi(\tfrac{1}{2}w_n)$, ..., $\phi'''(\tfrac{1}{2}w_n)$, whose exact form will not be needed, and
$$\psi''(0,w_n) = \frac{\partial^2\psi(u,w_n)}{\partial u^2}\bigg|_{u=0}$$

Thus
$$f_n(w_n) = \frac{n(n-1)\exp[(n-2)\psi(0,w_n)][\phi(\frac{1}{2}w_n)]^2}{[-\psi''(0,w_n)]^{\frac{1}{2}}}\int_{-\infty}^{\infty}[1 + A_2(w_n)t^2 + \dots]\exp\left[-\frac{(n-2)t^2}{2}\right]dt. \tag{18}$$

An asymptotic expansion in inverse powers of $(n-2)$ follows if the integral may be evaluated formally term by term. That this is in fact permissible may be shown by an application of Watson's lemma (Watson, 1944, p. 236).

Thus $$f_n(w_n) \sim \frac{n(n-1)(2\pi)^{\frac{1}{2}}\exp[(n-2)\psi(0,w_n)][\phi(\frac{1}{2}w_n)]^2}{[-\psi''(0,w_n)]^{\frac{1}{2}}}\left\{\frac{1}{(n-2)^{\frac{1}{2}}} + \frac{A_2(w_n)}{(n-2)^{\frac{3}{2}}} + \dots\right.$$
$$\sim \frac{n^{\frac{3}{2}}(2\pi)^{\frac{1}{2}}[\phi(\frac{1}{2}w_n)]^2}{[\phi'(-\frac{1}{2}w_n)]^{\frac{1}{2}}}[\Phi(\tfrac{1}{2}w_n) - \Phi(-\tfrac{1}{2}w_n)]^{n-\frac{3}{2}}\left\{1 + \frac{A_2(w_n)}{n-2} + \dots\right\}. \tag{19}$$

To obtain the last expression, the definition (15) of $\psi(u, w_n)$ has been used.

Put $$\Psi(x) = \Phi(x) - \Phi(-x). \tag{20}$$

Then the first term of the expansion (19) is
$$\frac{n^{\frac{3}{2}}(2\pi)^{\frac{1}{2}}[\phi(\frac{1}{2}w_n)]^2}{[\phi'(-\frac{1}{2}w_n)]^{\frac{1}{2}}}[\Psi(\tfrac{1}{2}w_n)]^{n-\frac{3}{2}}. \tag{21}$$

Now it can be shown that the function $A_2(w_n)$ is negative for the normal and many other distributions. A better approximation than (21) is therefore
$$\frac{c_n[\phi(\frac{1}{2}w_n)]^2[\Psi(\frac{1}{2}w_n)]^{n-\frac{3}{2}}}{[\phi'(-\frac{1}{2}w_n)]^{\frac{1}{2}}}, \tag{22}$$
where in this expression the constant c_n is to be chosen to make the integral from zero to infinity equal to unity.

In § 4 some numerical results are given comparing (22) with the exact distribution, for the case when the basic distribution is normal.

4. Numerical comparison with exact values

In this section the approximate results obtained above are compared with exact values for the normal distribution.

Figs. 1–4 give the frequency functions of range for $n = 20$ and $n = 50$ when the basic distribution is normal with unit standard deviation. The exact distribution was obtained from E. S. Pearson & H. O. Hartley's table (1942) for $n = 20$, and by numerical integration for $n = 50$.

Table 1 gives the exact values of mean, standard deviation, β_1 and β_2 of range, compared with the values from the formulae (8)–(11) of § 2, and with numerically computed values from the approximation of § 3 and from Elfving's approximation mentioned in § 1. The exact values, based on the work of Tippett (1925) and E. S. Pearson (1926), have been taken from *Tables for Statisticians and Biometricians*, Part II (K. Pearson, 1931, p. cxvii).

Table 1. *Constants of the range distribution in samples from a normal distribution of unit standard deviation*

n	Exact values				Approx. of § 2			
	Mean	S.D.	β_1	β_2	Mean	S.D.	β_1	β_2
10	3·08	0·80	0·16	3·22	3·22	1·03	0·81	4·2
20	3·73	0·73	0·16	3·26	3·85	0·88	0·81	4·2
50	4·50	0·65	0·19	3·34	4·58	0·75	0·81	4·2
100	5·02	0·60	0·22	3·39	5·09	0·68	0·81	4·2
500	6·07	0·52	0·28	3·50	6·12	0·57	0·81	4·2
1000	6·48	0·50	0·31	3·54	6·52	0·54	0·81	4·2

n	Approx. of § 3				Elfving's approx.			
	Mean	S.D.	β_1	β_2	Mean	S.D.	β_1	β_2
20	3·83	0·78	0·21	3·42	3·76	0·77	0·07	3·08
50	4·58	0·69	0·24	3·33	4·52	0·67	0·13	3·18

Figs. 1–4 show that the steepest descents approximation is more accurate than the approximation of § 2, although even for $n = 20$ the latter gives the ordinates of the range distribution fairly closely in the centre of the distribution but not in the tails.

This is confirmed by the moments given in the table. The Bessel function approximation of § 2 gives the mean range reasonably accurately, the standard deviation less accurately, while the β coefficients are very different (for the range of values of n considered) from the exact values. This is in accordance with the results of Fisher & Tippett, who showed that for the distribution of extremes the limiting β values are not reached until n is about 10^{12}.

The steepest descents approximation of § 3, and Elfving's approximation mentioned in § 1 are both more accurate. For $n = 50$ Elfving's approximation gives the mean range to

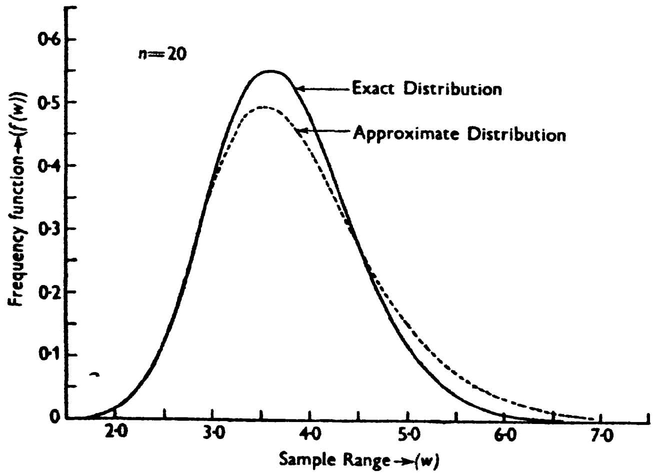


Fig. 1. Exact form compared with Bessel function approximation of § 2.

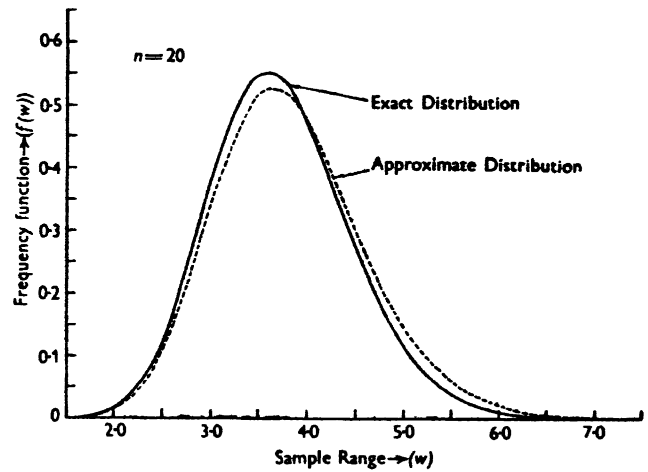


Fig. 2. Exact form compared with steepest descents approximation of § 3.

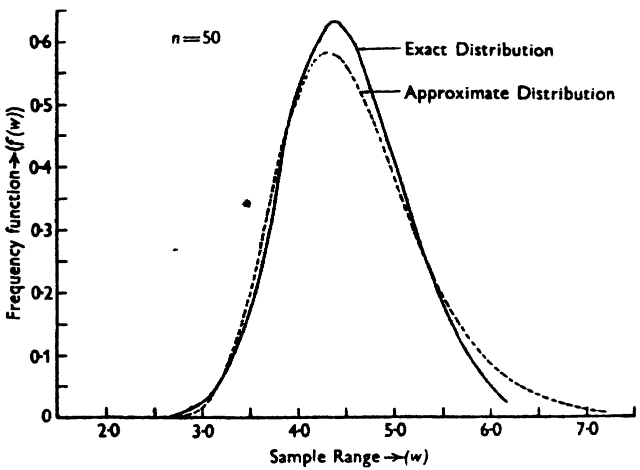


Fig. 3. Exact form compared with Bessel function approximation of § 2.

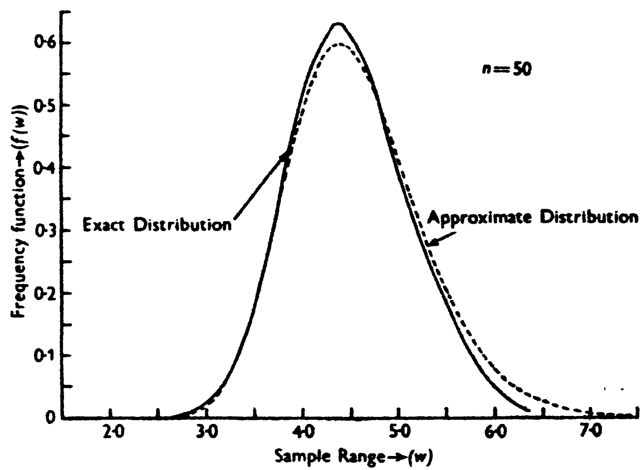


Fig. 4. Exact form compared with steepest descents approximation of § 3.

Fig. 1–4. Distribution of range in samples from normal population of unit standard deviation.

½ % and the standard deviation of range to 3 %. The steepest descents approximation is less accurate, the corresponding figures being 1½ and 6 %. The β coefficients appear to be given slightly more accurately by the steepest descents approximation.

The properties of the three approximations may be summarized as follows for the case when the basic distribution is normal.

Elfving's approximation is the most accurate with the steepest descents approximation second.

The Bessel function approximation gives fairly accurate estimates of the ordinates of the range distribution near the mode, but does not reproduce the ordinates in the tails until n is very large indeed.

The disadvantage of Elfving's method is that it involves a non-linear transformation of the range. This makes it unlikely that it can be used algebraically to simplify mathematical problems involving the distribution of range.

I wish to thank the Director of Research, Wool Industries Research Association, for permission to publish this paper, and Dr H. E. Daniels for his interest in the work.

REFERENCES

Elfving, G. (1947). *Biometrika*, **34**, 111.
Fisher, R. A. & Tippett, L. H. C. (1928). *Proc. Camb. Phil. Soc.* **24**, 180.
Gumbel, E. J. (1947). *Ann. Math. Statist.* **18**, 384.
Pearson, E. S. (1926). *Biometrika*, **18**, 173.
Pearson, E. S. & Hartley, H. O. (1942). *Biometrika*, **33**, 301.
Pearson, K. (1931). *Tables for Statisticians and Biometricians*, Part II. Cambridge University Press.
Tippett, L. H. C. (1925). *Biometrika*, **17**, 364.
Watson, G. N. (1944). *Theory of Bessel Functions*, 2nd ed. Cambridge University Press.

ON THE ROLE OF VARIABLE GENERATION TIME IN THE DEVELOPMENT OF A STOCHASTIC BIRTH PROCESS

By DAVID G. KENDALL, *Magdalen College, Oxford*

1. **Introduction.** The 'birth-and-death' process introduced by W. Feller (1939) provides a mathematical description of the growth of a population under the influence of very much simplified laws of reproduction and mortality. The population size n is considered as a random function of the time t, and the equations governing its development are derived from the following assumptions:

(i) The value of n at some initial time ($t = 0$) is supposed given. In the simplest case $n(0) = 1$.

(ii) It is assumed that the subpopulations stemming from two co-existing individuals will develop in complete independence of one another.

(iii) The risks of mortality and reproduction are supposed to be the same for each member of the population.

(iv) An individual known to be alive at time t has a chance asymptotically equal to λdt of reproducing itself, and a chance asymptotically equal to μdt of dying during the subsequent elementary time interval of length dt. Reproduction is here taken to imply binary subdivision, and so to result in the addition of just one member to the population.

(v) The chances of reproduction and mortality, described at (iv) above, are supposed to be completely independent of the previous history of the individual, and in particular to be independent of the time which has elapsed since its own 'birth' in some earlier subdivision.

(vi) The birth and death rates λ and μ are supposed to be independent of the epoch t.

Assumption (v) implies that the stochastic dependence of n upon t can be described with the aid of a discontinuous Markoff process,* and, in fact, it is easily seen that the functions

$$P_n(t) \equiv \text{Probability}\{n(t) = n \mid n(0) = 1\},$$

which completely determine the structure of the process, must satisfy the set of differential-difference equations

$$\left.\begin{aligned} \frac{d}{dt}P_n(t) &= (n+1)\mu P_{n+1}(t) + (n-1)\lambda P_{n-1}(t) - n(\lambda+\mu)P_n(t) \quad (n \geq 1), \\ \frac{d}{dt}P_0(t) &= \mu P_1(t). \end{aligned}\right\} \tag{1}$$

Feller's equations (1) were solved by C. Palm,† the mean and variance of n as functions of t having been already given in Feller's paper.

The idealized population whose growth is described by the equations (1) is, of course, rather far removed from reality, and it is therefore of some interest to consider how far any

* The development of a system is said to follow a Markoff process if its state at any time t can be described by the value of a random time-dependent variable $X(t)$ with the following property: *let the value of $X(t_0)$ be known; then if $t_1 > t_0$, the conditional distribution of the random variable $X(t_1)$ is in no way affected if the value of $X(t)$ is also given for any $t < t_0$.* Further details and some references will be found in my recent review (Kendall, 1947).

† Palm's formulae are quoted by N. Arley & V. Borchsenius (1945). See also M. S. Bartlett (1947) and D. G. Kendall (1948*a*).

relaxation is possible in the assumptions (i)–(vi). Of these the first two and the last two are the ones urgently requiring consideration; progress is most likely to be made by attacking them independently, and a good deal of work has already been done in this respect. If (ii) is retained, (i) is easily relaxed, for the $n(0)$ initial individuals then generate independent subpopulations each commencing with a single member, and it is therefore sufficient to raise the generating function for the $\{P_n(t)\}$ to the power $n(0)$. An interesting extension of (ii) has been considered by Feller, who discusses the birth-and-death process for which the rates λ and μ, instead of being constants, are linearly dependent on the instantaneous population size n. This model, the equations for which are as yet unsolved, corresponds to a population growing according to the logistic law of Pearl, Verhulst and Reed in the parallel deterministic theory. In a recent paper (Kendall, 1948*b*) I have given an account of the birth-and-death process in which the birth- and death-rates λ and μ, instead of being constants as in (vi), can be any desired functions of the epoch t; the most interesting example is perhaps that in which λ and μ are periodic functions of t.

The one important assumption which has not yet been relaxed is that of the Markoff property (v). This implies, for example, that in the absence of mortality an individual 'born' at time t will itself undergo subdivision at a time $t+\tau$, where the *generation time* τ has the distribution

$$e^{-\lambda\tau}\lambda\, d\tau \quad (0<\tau<\infty). \tag{2}$$

This is, of course, very different from the distributions of generation time actually observed; not only for man, but also for such elementary organisms as bacteria, the observed distribution usually possesses a pronounced non-zero mode, and the other extreme assumption of a *fixed* generation time $\tau = \tau_0$ (implying an exact doubling of the population at regular intervals) might seem to be more realistic.

Now the distribution (2) would assert that τ is a multiple of a χ^2-variate having two degrees of freedom, and this suggests that it would be worth while examining a modified process in which τ is distributed as the multiple of a χ^2_{2k}, where k is an integer greater than unity. The present paper is devoted to a development of this idea, and to simplify the analysis attention will here be confined to purely reproductive processes ($\mu = 0$). Before proceeding to details it is of interest to note what values of k are likely to be of practical relevance.

The most extensive information on the distribution of generation times for bacteria appears to be that communicated by C. D. Kelly & Otto Rahn (1932). They maintained bacteria in a 'warm stage' at 30° C., and measured (by continuous microscopic observation) the fission times for the second, third and fourth generations descended from each of a large number of individual cells. Their results concerning *Bacterium aerogenes* may be quoted in illustration. Observations were made, for this bacterium, on nine different days, the number of generation times measured per day varying from 30 to 126; the results for the three largest sets (each consisting of more than a hundred observations) are shown in histogram form in the accompanying diagram (Fig. 1). The sets relating to the several days have not been pooled because it is evident that there was a day-to-day variation in the mean generation time.

It will be seen at once that the assumption of a χ^2_{2k} distribution (with a change of scale to allow for the correct mean time) is at least not an unreasonable one, and that it is greatly superior to either of the alternatives previously available (which correspond to $k = 1$ and $k = \infty$). While it will certainly be of interest to examine the adequacy of the χ^2_{2k} hypothesis in more detail on another occasion, it is enough for the moment to know that it provides

a way of introducing into a model of population growth a law of variation of generation time which bears a general resemblance to that obtaining in reality.

It will be useful for the further developments of this paper to have in mind a rough estimate of the parameter k derived from the above data. Such an estimate can conveniently be based

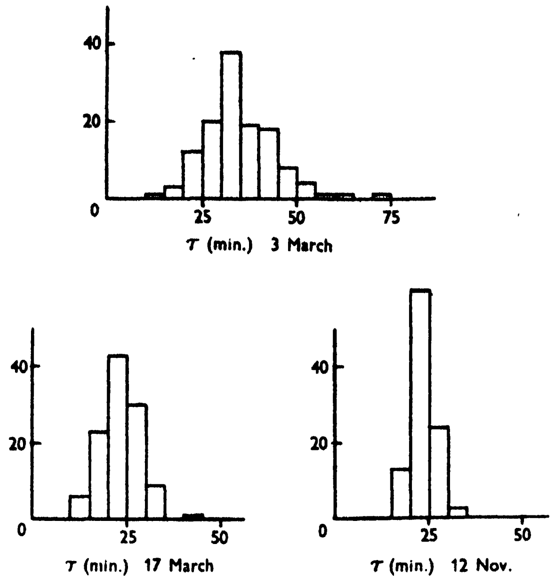


Fig. 1

Table 1. Bacterium aerogenes *at* 30° C. *Preliminary analysis of the observed generation times*, τ

(Data of Kelly and Rahn)

Date	No. of observations	Geometric mean of τ (min.)	'Coefficient of variation' (%)	Estimate of k
17 Feb.	44	38·5	17·6	33
24 Feb.	60	34·8	22·8	20
2 Mar.	84	28·9	27·7	14
3 Mar.	126	33·5	27·2	14
6 Mar.	84	29·6	23·0	19
10 Mar.	93	32·4	30·1	12
17 Mar.	112	22·6	23·9	18
12 Nov.	100	23·1	14·6	47
14 Nov.	30	24·9	24·7	17

on the observed variance of the natural logarithm of the generation time, which on the χ^2_{2k} hypothesis will have the expected value*

$$\kappa_2(\log\tau) = 1/(k-\tfrac{1}{2}), \quad \text{approximately.}$$

It is worth noting that the observed standard deviation of the logarithm of the generation time τ is roughly the same thing as the coefficient of variation of τ—it has been listed as such in Table 1. The observed variances of $\log\tau$, when analysed by M. S. Bartlett's well-known

* See, for example, M. S. Bartlett & D. G. Kendall (1946). (The error involved in this approximation is less than 1% for $k > 3$.)

test for homogeneity, were found to be significantly discordant ($\chi_8^2 = 62$), and no attempt has been made, therefore, to pool the nine estimates of k. For present purposes only the order of magnitude of k is required, and the value

$$k = 20$$

is evidently satisfactory, though, of course, there is no reason to suppose that it will be appropriate to describe the growth of any organism other than the one example discussed here.

2. **Specification of the multiple-phase birth process. The fundamental equations.** The process to be considered may be specified as follows. When a new individual is 'born', it passes through a series of phases, k in number, and only after it has attained the kth phase can it undergo subdivision. The lifetime in each phase is assumed to follow the law of distribution

$$e^{-k\lambda T}\, k\lambda\, dT \quad (0 < T < \infty),$$

the several lifetimes being independent, and the incident terminating the life of the individual in the kth phase being (i) its death and (ii) the birth, simultaneously, of two individuals who commence their existence in the first phase at that instant. Of course, the formal treatment would not in any way be altered if instead it were assumed that an individual terminates its existence in the kth phase by giving birth to one individual in the first phase, and simultaneously returning to the latter itself.

The multiple-phase birth process possesses the Markoff property, provided that its development is described by the *vector* variate $\mathbf{n}$, the k components of which enumerate the individuals existing in each of the k phases. If $\mathbf{n} \equiv (n_1, n_2, \dots, n_k)$, and

$$n = n_1 + n_2 + \dots + n_k,$$

then n, the total population size irrespective of phase, describes the growth of a birth process in which the generation time τ has the distribution of

$$\frac{1}{2k\lambda}\chi_{2k}^2.$$

When $k = 1$, the multiple-phase process is identical with the simple birth process discussed by Feller (i.e. that governed by equations (1), with $\mu = 0$), while if k is allowed to tend to infinity, one obtains the deterministic birth process with a fixed generation time $\tau_0 = 1/\lambda$. The multiple-phase process thus includes both the models already mentioned, and bridges the gap between them. Incidentally its study may throw some light on the effect of the 'Markoff' assumption in calculations of this sort. In this context, reference may be made to the general remarks in my review article (1947).

It will be supposed that initially the population consists of a single individual in the first phase. (Because of a well-known property of the distribution (2), it makes no difference whether this individual was born at the time $t = 0$ or at some earlier instant.) The course of events leading up to the first subdivision in what might be a typical case is shown in the accompanying diagram (Fig. 2). The stochastic development of the process will be fully described once the function

$$P\begin{pmatrix} n_k \\ \vdots \\ n_2 \\ n_1 \end{pmatrix},$$

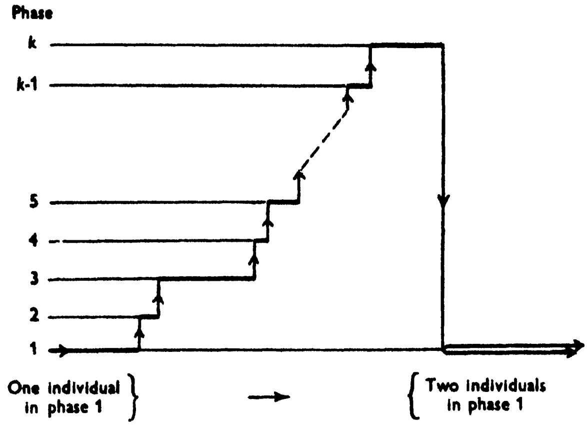


Fig. 2

is known; this is, the probability that at time t there will be n_1 individuals in the first phase, n_2 in the second phase, and so on. The differential-difference equations which correspond here to Feller's equations (1) can be written most concisely if one adopts the convention that $P \equiv 0$ whenever any of the n_i are negative. It will then readily be seen that

$$\frac{d}{dt}P\begin{pmatrix} n_k \\ \vdots \\ n_2, & t \\ n_1 \end{pmatrix} = (n_1+1)k\lambda P\begin{pmatrix} n_k \\ \vdots \\ n_2-1, & t \\ n_1+1 \end{pmatrix} + \dots$$

$$+ (n_{k-1}+1)k\lambda P\begin{pmatrix} n_k-1 \\ n_{k-1}+1, & t \\ \vdots \\ n_1 \end{pmatrix} + (n_k+1)k\lambda P\begin{pmatrix} n_k+1 \\ \vdots \\ n_2, & t \\ n_1-2 \end{pmatrix}$$

$$-(n_1+n_2+\dots+n_k)k\lambda P\begin{pmatrix} n_k \\ \vdots \\ n_2, & t \\ n_1 \end{pmatrix}, \tag{3}$$

and if one introduces the generating function

$$\phi\begin{pmatrix} z_k \\ \vdots \\ z_2, & t \\ z_1 \end{pmatrix} = \Sigma z_1^{n_1} z_2^{n_2} \dots z_k^{n_k} P\begin{pmatrix} n_k \\ \vdots \\ n_2, & t \\ n_1 \end{pmatrix}, \tag{4}$$

it will follow that this must satisfy the partial differential equation

$$\frac{1}{k\lambda}\frac{\partial\phi}{\partial t} = \left(z_2\frac{\partial\phi}{\partial z_1} + z_3\frac{\partial\phi}{\partial z_2} + \dots + z_k\frac{\partial\phi}{\partial z_{k-1}}\right) + z_1^2\frac{\partial\phi}{\partial z_k} - \sum_{i=1}^{k} z_i\frac{\partial\phi}{\partial z_i}, \tag{5}$$

the associated boundary condition being of course

$$\phi \equiv z_1 \quad \text{when} \quad t = 0. \tag{6}$$

The partial differential equation (5) is of the standard Lagrangian form, the auxiliary equations being

$$\frac{d\phi}{0} = \frac{k\lambda\, dt}{1} = \frac{dz_1}{z_1-z_2} = \frac{dz_2}{z_2-z_3} = \dots = \frac{dz_{k-1}}{z_{k-1}-z_k} = \frac{dz_k}{z_k-z_1^2}. \tag{7}$$

To solve these, it is convenient to introduce a new time scale defined by $k\lambda t = \theta$, and to write $X_i = e^{-\theta}z_i$; the equations (7) then become

$$X'_j = -X_{j+1} \quad (j = 1, 2, \ldots, k-1), \qquad X'_k = -e^{\theta}X_1^2,$$

and so
$$X_i = \left(-\frac{d}{d\theta}\right)^{i-1} X(\theta), \tag{8}$$

where $X(x)$ satisfies the ordinary differential equation

$$\left(-\frac{d}{dx}\right)^k X(x) = e^x\{X(x)\}^2. \tag{9}$$

Let the general solution of (9) be

$$X(x) = F(x; c_1, c_2, \ldots, c_k),$$

where the arbitrary constants $\{c_i\}$ are so chosen that

$$c_i = F^{(i-1)}(0; c_1, c_2, \ldots, c_k) \quad (1 \leqslant i \leqslant k). \tag{10}$$

Then
$$X_i = (-)^{i-1}F^{(i-1)}(\theta; c_1, c_2, \ldots, c_k), \tag{11}$$

and if these equations when solved for $\{c_j\}$ take the form

$$c_j = G_j(\theta; X_1, X_2, \ldots, X_k), \tag{12}$$

the general solution to (5) will be*

$$\phi(z_1, z_2, \ldots, z_k; t) \equiv \Phi(G_1, G_2, \ldots, G_k), \tag{13}$$

where Φ is an arbitrary function of its k arguments.

The unknown function Φ must now be identified with the aid of the boundary condition (6). To this end, put $\theta = 0$ in (11) and (12) and compare with (10); it will then be seen that

$$G_j(0; u_1, u_2, \ldots, u_k) \equiv (-)^{j-1}u_j.$$

From the boundary condition (6), however, it follows that

$$\begin{aligned} z_1 &\equiv \phi(z_1, z_2, \ldots, z_k; 0) \equiv \Phi\{\ldots, G_j(0; z_1, z_2, \ldots, z_k), \ldots\} \\ &\equiv \Phi(z_1, -z_2, \ldots, (-)^{k-1}z_k), \end{aligned}$$

and so the function Φ must be given by

$$\Phi(G_1, G_2, \ldots, G_k) \equiv G_1.$$

The generating function ϕ is thus $G_1(\theta; X_1, X_2, \ldots, X_k)$, and this is the value of c_1, i.e. of $F(0; c_1, c_2, \ldots, c_k)$, when the $\{c_j\}$ are to be determined from the k equations

$$z_i e^{-\theta} = (-1)^{i-1}F^{(i-1)}(\theta; c_1, c_2, \ldots, c_k).$$

Expressing this result in different words: $\phi(z_1, z_2, \ldots, z_k; t)$ is equal to the value of $X(0)$, when $X(x)$ satisfies the equation (9) together with the boundary conditions

$$X(\theta) = z_1e^{-\theta}, \quad X'(\theta) = -z_2e^{-\theta}, \quad \ldots, \quad X^{(k-1)}(\theta) = (-)^{k-1}z_ke^{-\theta}.$$

Before working out an example in detail, it is convenient to throw the solution into a more familiar form by writing $x = \theta - u$ and $X(x) = e^{-\theta}Z(u)$. It will then be seen that *the generating function ϕ for the multiple-phase birth process is equal to $e^{-k\lambda t}Z(k\lambda t)$, where the function $Z(u)$ is determined by the differential equation*

$$\left(\frac{d}{du}\right)^k Z(u) = e^{-u}\{Z(u)\}^2, \tag{14}$$

and the boundary conditions
$$Z^{(i)}(0) = z_{i+1} \quad (0 \leqslant i \leqslant k-1). \tag{15}$$

* The arguments of ϕ will henceforth, for typographical convenience, be written horizontally.

This is a perfectly straightforward one-point boundary problem, although unfortunately the equation (14) seems to be intractable for values of k greater than unity. A determination of the P-functions for the multiple-phase process is therefore not practicable; it is, however, possible to discuss the mean and variance of the distribution, and its approximate normality for large k, without solving the equation (14). These matters will be considered in the following sections of the paper; for the moment it is of interest to leave the general argument and examine the solution of the equation (14) in the elementary case when $k = 1$.

The equation is then $Z' = e^{-u}Z^2$, with the single boundary condition $Z(0) = z$. The general solution is

$$\frac{1}{Z} - e^{-u} = C,$$

and on identifying the arbitrary constant this becomes

$$Z(u) = \frac{z}{1 - z(1 - e^{-u})}.$$

Accordingly the generating function is

$$\phi = \frac{ze^{-\lambda t}}{1 - z(1 - e^{-\lambda t})}$$

and
$$P(n, t) = e^{-\lambda t}(1 - e^{-\lambda t})^{n-1} \quad (n \geq 1),$$

in agreement with Feller's original calculation (1939, equation (17) with $N = 1$). For future reference it will be convenient to note here the mean and variance of this distribution, as already given by Feller; they are

$$E(n) = e^{\lambda t} \quad \text{and} \quad \operatorname{Var}(n) = e^{\lambda t}(e^{\lambda t} - 1). \tag{16}$$

3. **The mean growth of the process.** It is possible to deduce the differential equations satisfied by the expected values of the $\{n_j\}$ from the fundamental equation (14), but it is easier to derive them independently; in fact, simple considerations of continuity give at once

$$\frac{d\nu_1}{d\theta} = 2\nu_k - \nu_1,$$

and
$$\frac{d\nu_j}{d\theta} = \nu_{j-1} - \nu_j \quad (1 < j \leq k), \tag{17}$$

where ν_j is the expected value of n_j. If now one writes ρ_j for $\nu_j e^{\theta}$ these equations can readily be seen to be equivalent to

$$\rho_j = \left(\frac{d}{d\theta}\right)^{k-j} \rho,$$

where
$$\left(\frac{d}{d\theta}\right)^{k} \rho = 2\rho.$$

Thus if ω is the primitive kth root of unity, $\exp(2\pi i/k)$, then

$$\rho = \sum_{r=0}^{k-1} A_r \exp(2^{1/k}\omega^r\theta)$$

(where the quantities A_r are as yet undetermined constants), and

$$\rho_j = 2^{1-j/k} \sum_{r=0}^{k-1} A_r \omega^{-jr} \exp(2^{1/k}\omega^r\theta).$$

Now initially, when both t and θ are zero, $\rho_1 = 1$ and $\rho_j = 0\ (1 < j \leq k)$; the $\{A_r\}$ must therefore have the values

$$A_r = \frac{1}{2k} 2^{1/k}\omega^r.$$

Accordingly, the expected number of individuals in the jth phase at time t is

$$E(n_j) = \frac{1}{k} 2^{(1-j)/k} \sum_{r=0}^{k-1} \omega^{-(j-1)r} \exp\{(2^{1/k}\omega^r - 1) k\lambda t\}, \tag{18}$$

while the expected total population size at the same time is

$$E(n) = \frac{1}{2k} \sum_{r=0}^{k-1} \frac{2^{1/k}\omega^r}{2^{1/k}\omega^r - 1} \exp\{(2^{1/k}\omega^r - 1) k\lambda t\}, \tag{19}$$

$$= e^{-k\lambda t} \sum_{m=0}^{\infty} 2^m \left\{ \frac{(k\lambda t)^{mk}}{(mk)!} + \frac{(k\lambda t)^{mk+1}}{(mk+1)!} + \dots + \frac{(k\lambda t)^{mk+k-1}}{(mk+k-1)!} \right\}. \tag{20}$$

In the last formula, the jth term in each bracket is the contribution from the jth phase.

These formulae possess two limiting forms of practical interest. The first describes (for fixed k) the behaviour of the mean population size for large values of the time t. The other limiting form is obtained by fixing t and allowing k to approach infinity (so that the process approximates to the deterministic model for which the generation time has the fixed value $1/\lambda$). Care is required in interpreting the results because the two limiting processes do *not* in general commute.

For large values of t (k remaining fixed) the dominant term in (18) is always the first. Indeed, since

$$2^{1/k} \cos \frac{2\pi}{k} - 1 < 0 \quad \text{when} \quad 2 \leqslant k \leqslant 28,$$

the remaining terms represent damped oscillations for these values of k, while for larger values of k they do not themselves tend to zero but are still of a lower order than the first term. Since

$$\lim_{k \to \infty} k(x^{1/k} - 1) = \log x,$$

it is convenient to write

$$\alpha_k = k(2^{1/k} - 1) \quad (\to \log 2 = 0{\cdot}693, \text{ as } k \to \infty).$$

The expected population sizes for large t are then given asymptotically by

$$E(n_j) \simeq \frac{2^{-(j-1)/k}}{k} e^{\alpha_k \lambda t}, \tag{21}$$

and

$$E(n) \simeq \frac{2^{1/k}}{2\alpha_k} e^{\alpha_k \lambda t}. \tag{22}$$

It is rather curious that the coefficient of $e^{\alpha_k \lambda t}$ in (22) has the limit 0·721 as k approaches infinity, and not unity as one might have expected. This is an example of the non-commuting character of the two limiting processes.

The dependence of α_k upon k is shown by the following short table. As far as the ultimate rate of growth of the mean population size is concerned, the variability of the generation time evidently ceases to have much effect after k exceeds a value of about 35 (this is the value of k for which α_k differs from log 2 by 1 %):

$k =$ 1	2	3	4	5	10	15	20	25	30	∞
$\alpha_k =$ 1·000	0·828	0·780	0·757	0·744	0·718	0·709	0·705	0·703	0·701	0·693

It is more difficult to discuss the behaviour of $E(n)$ when k tends to infinity and t remains fixed. From (20) it follows that $E(n)$ is equal to the expected value of $2^{[N/k]}$, when N is a Poisson variable of mean value $k\lambda t$, and $[x]$ is used to denote the integer part of x. It can be shown from this that

$$\lim_{k \to \infty} E(n) = 2^{[\lambda t]} \tag{23}$$

when λt is not an integer;

while
$$\lim_{k\to\infty} E(n) = \tfrac{3}{4}\,.\,2^{\lambda t} \tag{23a}$$
when λt is an integer. The limit of $E(n)$ as k approaches infinity is thus a discontinuous function of the time t, whose value at a point of discontinuity is equal to the arithmetic mean of the left- and right-hand limits. The details of this calculation, though elementary, are a little tiresome, and are perhaps not worth reproducing here. Formula (23) of course expresses the fact that in the limit as k tends to infinity the multiple-phase model coincides with the deterministic model already mentioned.

It will be noticed from (21) that for fixed k and large values of t, the expected numbers of individuals in each of the several phases will be proportional to
$$1, \quad 2^{-1/k}, \quad 2^{-2/k}, \quad \ldots, \quad 2^{-(k-1)/k}. \tag{24}$$
When $k = 2$, it is possible to express the formulae for the expected population sizes in closed form:
$$E(n_1) = e^{-2\lambda t}\cosh(\lambda t\, 2\sqrt{2}) \quad \text{and} \quad E(n_2) = \frac{1}{\sqrt{2}}e^{-2\lambda t}\sinh(\lambda t\, 2\sqrt{2}). \tag{25}$$

4. **The differential equation for the cumulant-generating function.** The cumulant-generating function for the total population size is
$$\log E(e^{n\alpha}) = K(\alpha,\theta) = \log\phi(e^{\alpha}, e^{\alpha}, \ldots, e^{\alpha}; t) = -\theta + \log Z(\theta),$$
where the function $Z(\theta)$ is to be determined from the differential equation
$$\left(\frac{d}{d\theta}\right)^k Z(\theta) = e^{-\theta}\{Z(\theta)\}^2,$$
with the boundary conditions $Z^{(i)}(0) = e^{\alpha} \quad (0 \leqslant i \leqslant k-1)$.
If in these equations one writes $Z = e^{\theta+K}$, the boundary conditions become
$$\left(1+\frac{\partial}{\partial\theta}\right)^i e^{K(\alpha,\theta)} = e^{\alpha}, \quad \text{when} \quad \theta = 0 \quad (0 \leqslant i \leqslant k-1),$$
or
$$e^{K(\alpha,0)} = e^{\alpha} \quad \text{and} \quad \left(\frac{\partial}{\partial\theta}\right)^i e^{K(\alpha,0)} = 0 \quad (1 \leqslant i \leqslant k-1),$$
from which it is easily seen that
$$K(\alpha,0) = \alpha \quad \text{and} \quad \left(\frac{\partial}{\partial\theta}\right)^i K(\alpha,0) = 0 \quad (1 \leqslant i \leqslant k-1), \tag{26}$$
while the differential equation for $K(\alpha,\theta)$ which is to be associated with these boundary conditions is
$$\left(1+\frac{\partial}{\partial\theta}\right)^k e^{K(\alpha,\theta)} = e^{2K(\alpha,\theta)}. \tag{27}$$
For some purposes it is convenient to write $\theta = kT$, so that $T = \lambda t$, and then
$$\left(1+\frac{1}{k}\frac{\partial}{\partial T}\right)^k e^{K} = e^{2K}, \tag{28}$$
while
$$K = \alpha \quad \text{and} \quad \left(\frac{\partial}{\partial T}\right)^i K = 0, \quad \text{when} \quad T = 0 \quad (1 \leqslant i \leqslant k-1). \tag{29}$$

The two most important special cases are, of course, (i) $k = 1$, and (ii) $k \to \infty$. In case (i) let $e^K \equiv M$; then the differential equation becomes
$$\frac{\partial M}{\partial T} = M(M-1), \quad \text{with} \quad M(0) = e^{\alpha},$$
which is easily found to have the solution
$$M = \frac{1}{1-e^{T}(1-e^{-\alpha})},$$

in agreement with the results of Feller mentioned at the end of § 2. On the other hand, in case (ii), as $k \to \infty$ the operator
$$\left(1+\frac{1}{k}\frac{\partial}{\partial T}\right)^k$$
formally approaches the limit
$$\exp\left(\frac{\partial}{\partial T}\right),$$
and so the equation (28) takes the form
$$K(\alpha, T+1) = 2K(\alpha, T)$$
with the initial conditions
$$K(\alpha, 0) = \alpha \quad \text{and} \quad \left(\frac{\partial}{\partial T}\right)^i K(\alpha, 0) = 0 \quad (\text{all } i \geqslant 1).$$
The solution is
$$K(\alpha, T) = \alpha 2^{[T]},$$
in agreement with the results for the deterministic model.

5. **The variance of n.** From the equation satisfied by the cumulant-generating function one can readily obtain an equation to determine the variance of the population size n. Since
$$e^K = 1 + \alpha\kappa_1 + \tfrac{1}{2}\alpha^2(\kappa_2 + \kappa_1^2) + \dots,$$
where the cumulants κ_1, κ_2, etc., are functions of θ, it follows that
$$\left(1+\frac{d}{d\theta}\right)^k \kappa_1 = 2\kappa_1, \tag{30}$$
and
$$\left(1+\frac{d}{d\theta}\right)^k (\kappa_2+\kappa_1^2) = 2(\kappa_2+\kappa_1^2) + 2\kappa_1^2. \tag{31}$$
The first of these equations provides an alternative starting-point for the investigation of the mean value of n, given in § 3, while the second is the required formula determining the variance. It is to be combined with the boundary conditions
$$\left(\frac{d}{d\theta}\right)^i \kappa_2(0) = 0 \quad (0 \leqslant i \leqslant k-1).$$
In principle the equation (31) could be solved explicitly, but fortunately there is little to be gained in carrying out this laborious task. It is enough to notice, first, that when the time t is large the dominating term in the complementary function will be a constant multiple of
$$e^{(2^{1/k}-1)\theta}, \tag{32}$$
while the dominating term in the particular integral is
$$\frac{2}{\left(1+\frac{2}{k}\alpha_k\right)^k - 2}\left\{\frac{2^{1/k}}{2\alpha_k}e^{(2^{1/k}-1)\theta}\right\}^2. \tag{33}$$
Thus, since (32) is of a smaller order than (33) when θ is large, it follows that
$$\text{Variance}(n) \simeq \left\{\frac{2}{\left(1+\frac{2\alpha_k}{k}\right)^k - 2} - 1\right\}\bar{n}^2, \quad = C_k\bar{n}^2, \tag{34}$$
as the time t tends to infinity. The values of C_k for the first few values of k are given below:

$k =$	1	2	3	4	5
$C_k =$	1·000	0·489	0·324	0·242	0·193

If k is allowed to approach infinity,

$$C_k \simeq \frac{2(\log 2)^2}{k}, \quad = \frac{0{\cdot}9609}{k};$$

the table shows that this asymptotic formula is true with surprising accuracy even for quite small values of k (the error when $k = 1$ being only 4 %). Accordingly the relation

$$\text{C. of V.}(n) \simeq \sqrt{\frac{2}{k}}\log 2, \tag{35}$$

for the coefficient of variation of the population size can be used for all values of k, provided that $\bar{n}$ is sufficiently large. In the following section an extension of (35) will be found, which can be used whenever *one* of k and $\bar{n}$ is large enough. Some values of (35) are tabulated below (although the first entry has been based on the exact value $C_1 = 1$).

C. *of* V. (n), *for large* $\bar{n}$

k	= 1	5	20	100
C. of V. (n)	= 100 %	44 %	22 %	10 %

6. The asymptotic form for the distribution of the population size when the parameter k tends to infinity. Once again it is convenient to write $\theta = kT$ and $e^K = M$, so that $T = \lambda t$; $M(\alpha, T)$ is then the moment-generating function for the population size n, and satisfies the differential equation

$$\left(1 + \frac{1}{k}\frac{\partial}{\partial T}\right)^k M = M^2, \tag{36}$$

with the boundary conditions

$$M = e^{\alpha} \quad \text{and} \quad \left(\frac{\partial}{\partial T}\right)^i M = 0 \quad (1 \leqslant i \leqslant k-1), \quad \text{when} \quad T = 0.$$

It was seen in the last section that for large T and k the coefficient of variation of n is of order $k^{-\frac{1}{2}}$; this suggests that it may be appropriate to consider the standardized variable

$$X \equiv k^{\frac{1}{2}}(n - 2^T), \tag{37}$$

which one might expect to have a non-trivial limiting distribution as k tends to infinity, for any fixed value of t. The variable X has the moment-generating function

$$E(e^{\beta X}) \equiv M_0(\beta, T) = \exp(-k^{\frac{1}{2}}\beta\, 2^T)\, M,$$

where M is the solution of (36) associated with the boundary conditions

$$M = \exp(\beta k^{\frac{1}{2}}) \quad \text{and} \quad \left(\frac{\partial}{\partial T}\right)^i M = 0 \quad (1 \leqslant i \leqslant k-1), \quad \text{when} \quad T = 0.$$

The problem is, therefore, to substitute

$$\exp(k^{\frac{1}{2}}\beta\, 2^T)\, M_0(\beta, T)$$

for M in (36), and then to find the asymptotic form (as k tends to infinity) of the solution satisfying the boundary conditions

$$M_0 = 1 \quad \text{and} \quad \left(\frac{\partial}{\partial T}\right)^i M_0 = \exp(\beta k^{\frac{1}{2}})\left(\frac{\partial}{\partial T}\right)^i \exp(-k^{\frac{1}{2}}\beta\, 2^T) \quad (1 \leqslant i \leqslant k-1), \quad \text{when} \quad T = 0. \tag{38}$$

The formal solution to this problem will now be given. A rigorous treatment would be preferable, but one does not immediately suggest itself; it therefore seems worth while pursuing the investigation by heuristic methods.

In the first place, it is convenient to write

$$\Lambda(x) \equiv e^{-x}\left(1+\frac{x}{k}\right)^k \equiv \sum_{m=0}^{\infty} f_m(k)\frac{x^m}{k^{m-1}}, \tag{39}$$

and then to write the fundamental equation (36) in the form

$$\Lambda\left(\frac{\partial}{\partial T}\right) M(T) = \{M(T-1)\}^2,$$

or, on expansion,

$$\sum_{m=0}^{\infty} \frac{f_m(k)}{k^{m-1}}\left(\frac{\partial}{\partial T}\right)^m M(T) = \{M(T-1)\}^2. \tag{40}$$

The first few coefficients $f_m(k)$ are

$$\begin{aligned} f_0(k) &= 1/k, & f_1(k) &= 0, \\ f_2(k) &= -\tfrac{1}{2}, & f_3(k) &= +\tfrac{1}{3}, \\ f_4(k) &= -\tfrac{1}{4}+\tfrac{1}{8}k, & f_5(k) &= +\tfrac{1}{5}-\tfrac{1}{6}k, \end{aligned}$$

and on considering the mode of formation of the general term of the expansion* it will be found that when $m = 2p+1$ the largest term in $f_m(k)$ is of order k^{p-1}, while when $m = 2p$ the largest term is

$$\frac{(-)^p}{2^p p!} k^{p-1}. \tag{41}$$

On the other hand, for all values of m the largest term in

$$\left(\frac{\partial}{\partial T}\right)^m \{\exp(k^{\frac{1}{2}}\beta\, 2^T) M_0(\beta, T)\},$$

as k tends to infinity (supposing M_0 and its derivatives to remain of finite order in these circumstances), is

$$k^{\frac{1}{2}m}(\beta\, 2^T \log 2)^m \exp(k^{\frac{1}{2}}\beta\, 2^T) M_0(\beta, T).$$

Thus if

$$\lim_{k\to\infty} M_0(\beta, T) = M_1(\beta, T),$$

it will follow that

$$\begin{aligned} \frac{\{M_1(\beta, T-1)\}^2}{M_1(\beta, T)} &= \sum_{p=0}^{\infty} \frac{(-)^p}{p!}\{\tfrac{1}{2}\beta^2(\log 2)^2\, 2^{2T}\}^p \\ &= \exp\{-\tfrac{1}{2}\beta^2(\log 2)^2\, 2^{2T}\}, \end{aligned} \tag{42}$$

and the most general solution of this functional equation (when T is an integer) is easily seen to be

$$M_1(\beta, T) = \exp\{\beta^2(\log 2)^2\, 2^{2T} + A\, 2^T\},$$

where the constant A is to be determined from the initial conditions. Now

$$M_1(\beta, 0) = M_0(\beta, 0) = 1,$$

and so when T is an integer,

$$M_1(\beta, T) = \exp\{\tfrac{1}{2}\beta^2\, 2(\log 2)^2\, 2^T(2^T-1)\}. \tag{43}$$

But this is the limiting form of the moment-generating function for the standardized random variable (37). Accordingly the above argument indicates that *as k tends to infinity the population size n is asymptotically normally distributed with the mean value*

$$\bar{n} = 2^{\lambda t} \tag{44}$$

and the variance

$$\frac{2(\log 2)^2}{k}\bar{n}(\bar{n}-1), \tag{45}$$

for every fixed integer value of λt.

* It is advisable to calculate the $f_m(k)$ from the successive expansion of $\exp\{\log\Lambda(x)\}$.

The rather unexpected agreement with formula (35) of § 5 will be noticed. It appears, in fact, that the approximate formula,

$$\text{C. of V.}\,(n) \simeq \log 2\left(\frac{2}{k}\right)^{\frac{1}{2}}\left(1-\frac{1}{\bar{n}}\right)^{\frac{1}{2}}, \tag{46}$$

is valid whenever *one* of k and $\bar{n}$ is sufficiently large. The limiting values of (46), for large $\bar{n}$, are illustrated by the second table in § 5.

The purely formal character of the preceding argument is clearly responsible for the difference between (44) and the earlier, precise, result (23a). It is to be expected that the result (45) may be similarly incomplete, and a more careful analysis would be of interest.

7. **The coefficient of variation of the population size for more general types of stochastic birth process.** It has been shown in § 5 that for the multiple-phase birth process

$$\text{C. of V.}\,(n) \simeq A_k/k^{\frac{1}{2}}, \quad \text{when} \quad \bar{n} \text{ is large},$$

where the coefficient A_k is equal to unity when $k = 1$ and has the limit 0·98 as k approaches infinity. On the other hand, it easily follows from the definition of the process that

$$\text{C. of V.}\,(\tau) = 1/k^{\frac{1}{2}}$$

(where τ, as usual, denotes the generation time). The similarity of the two results is striking, and their relationship will be made clearer by the following crude argument, a precise formulation of which is yet to be found. This gives reasons for supposing that the relation,

$$\text{C. of V.}\,(n) \simeq 0{\cdot}98\,\text{C. of V.}\,(\tau), \quad \text{when} \quad \bar{n} \text{ is large}, \tag{47}$$

is true for *all* stochastic birth processes of the general type discussed in § 1, *whatever the form of the distribution of generation time*, provided that C. of V. (τ) is small enough for the process to be equivalent, in regard to its mean growth, to the deterministic process for which

$$n = 2^{\lambda t},$$

where $1/\lambda$ is the mean generation time for the *actual* process. To cast the subsequent argument into a rigorous form it would be necessary to pay careful attention to the 'spreading' of generations; this will not be attempted here, and the conclusions should be regarded as tentative only.

Suppose then that t (and so also $\bar{n}$) is so large as to make it 'practically certain' that the first g generations have been completely established.*

Let $$\tau_{ij} \quad (i = 1, 2, \ldots, g-1;\; j = 1, 2, \ldots, 2^{i-1})$$

be the time which elapses between the 'birth' of an individual (to be identified by the suffix j) in the ith generation and its own later subdivision. Thus, for example, τ_{11} is the epoch at which the initial individual subdivides, to be replaced by the two individuals of the second generation which themselves subdivide at the epochs $\tau_{11}+\tau_{21}$ and $\tau_{11}+\tau_{22}$ respectively. When the gth generation has been established its 2^{g-1} members will continue to generate 2^{g-1} independent subpopulations during the intervals of time severally left to them until the final count is taken at the epoch t. If any one member of the gth generation is 'born' at the epoch u, it will have a time $(t-u)$ available for continued subdivision, and the size of the subpopulation which it generates will be (in the deterministic model)

$$2^{\lambda(t-u)}.$$

* One (artificial) way of making this statement precise would be to impose an upper bound to the possible values of the generation time. This, however, would exclude the multiple-phase process.

It is a principle of the present calculation that once the first-order fluctuations have been identified their coefficients can be calculated on the basis of the deterministic model. A fluctuation δu in u thus implies a fluctuation

$$-\lambda\,\delta u\, 2^{\lambda(t-u)}\log 2$$

in the *expected* size of the subpopulation developed at the epoch t, and on inserting the 'deterministic' value for u in the coefficient, this becomes

$$-\lambda\,\delta u\, 2^{\lambda t-g+1}\log 2.$$

Now δu is the sum of the fluctuations in the generation times for each of the specified individual's $(g-1)$ ancestors. If for the moment the process is imagined to develop strictly according to expectation once the gth generation has been established, it will be seen that the contribution of the fluctuation of τ_{ij} to the final population size n will be

$$-2^{g-1}\lambda\,\delta\tau_{ij}\,2^{\lambda t-g+1}\log 2,$$

multiplied by the fraction of individuals in the gth generation who possess (i,j) as an ancestor. This fraction is

$$1/2^{i-1},$$

and so

$$\frac{\partial n}{\partial\tau_{ij}} = -\lambda\, 2^{\lambda t-i+1}\log 2. \tag{48}$$

The total contribution to $\operatorname{Var}(n)$ from the fluctuations possible during the development of the first g generations is thus

$$\lambda^2(\log 2)^2\operatorname{Var}(\tau)\{1(2^{\lambda t})^2+2(2^{\lambda t-1})^2+2^2(2^{\lambda t-2})^2+\ldots+2^{g-2}(2^{\lambda t-g+2})^2\}$$
$$= 2\lambda^2(\log 2)^2\operatorname{Var}(\tau)\,2^{2\lambda t}\left\{1-\frac{1}{2^{g-1}}\right\}.$$

To this must be added the further contribution arising from fluctuations in the development of the $(g+1)$th and later generations. Since this is equivalent to the growth of 2^{g-1} independent populations during a time $\left(t-\dfrac{g-1}{\lambda}\right)$, the additional contribution to the variance of n will be

$$2^{g-1}\,2\lambda^2(\log 2)^2\operatorname{Var}(\tau)\,2^{2\lambda t-2g+2},$$

to the first order (adopting the usual iterative procedure). If now $\bar{n}$ is large enough to justify the use of a sufficiently large g, it will follow that the second contribution is negligible and that

$$\text{C. of V.}(n) \simeq 2^{\frac{1}{2}}\log 2\ \text{C. of V.}(\tau),$$

as required. Since $2^{\frac{1}{2}}\log 2 = 0{\cdot}98$, it is therefore suggested as a practical rule-of-thumb that *the coefficients of variation for the population size and the generation time are approximately equal.* Of course, if the initial population size N is greater than unity this estimate of C. of V. (n) must be divided by $\sqrt{N}$, since each of the N initial members will generate independently a subpopulation to which the result will apply.

SUMMARY

This paper presents a mathematical account of a stochastic birth process in which the generation time (the interval between 'birth' and 'parenthood' in the life of an individual) is distributed like a χ^2-variate with $2k$ degrees of freedom. When $k = 1$, the process reduces to the simple stochastic birth process in the form originally introduced by W. Feller (1939), while as k tends to infinity the process assumes a strictly deterministic form in which the

population undergoes an exact doubling at regular intervals. It is suggested that for intermediate values of k (say of the order of 20) this new 'multiple-phase' stochastic birth process represents a further step towards the construction of an adequate mathematical model of the growth of real populations of elementary organisms.

In order to show that the methods of the paper are relevant in an actual biological example, a brief discussion is given of the work of Kelly & Rahn (1932) on the distribution of generation times for *Bacterium aerogenes*.

The paper concludes with a sketch of a more general argument suggesting that for a wide class of stochastic birth processes the coefficient of variation of the population size is ultimately approximately equal to the coefficient of variation of the generation time, when this is sufficiently small for the process to approximate to the deterministic form, the population being assumed to have developed from a single 'ancestor' in the absence of 'mortality'.

In conclusion, I should like to express my thanks to Mr D. J. Finney for kindly making available to me the computing facilities at his disposal, during the preparation of this paper.

REFERENCES

Arley, N. & Borchsenius, V. (1945). On the theory of infinite systems of differential equations and their application to the theory of stochastic processes and the perturbation theory of quantum mechanics. *Acta Math.* **76**, 261–322, esp. pp. 298–9.

Bartlett, M. S. (1947). *Stochastic Processes.* (Notes of a course given at the University of North Carolina in the Fall Quarter, 1946.) It is understood that copies of these notes are available on request.

Bartlett, M. S. & Kendall, D. G. (1946). The statistical analysis of variance-heterogeneity, and the logarithmic transformation. *J. R. Statist. Soc. Suppl.*, **8**, 128–38.

Feller, W. (1939). Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung. *Acta Biotheoretica*, **5**, 11–40.

Kelly, C. D. & Rahn, O. (1932). The growth rate of individual bacterial cells. *J. Bacteriol.* **23**, 147–53.

Kendall, D. G. (1947). A review of some recent work on discontinuous Markoff processes with applications to biology, physics, and actuarial science. *J. R. Statist. Soc.* **110**, 130–7.

Kendall, D. G. (1948*a*). On some modes of population growth leading to R. A. Fisher's logarithmic series distribution. *Biometrika*, **35**, 6–15.

Kendall, D. G. (1948*b*). On the generalized 'birth-and-death' process. *Ann. Math. Statist.* **19**, 1–15.

2×2 TABLES; THE POWER FUNCTION OF THE TEST ON A RANDOMIZED EXPERIMENT

By E. S. PEARSON and MAXINE MERRINGTON

1. Introductory

In his discussion of significance tests for 2×2 tables, Barnard (1947) has pointed out how data classified in the form of Table 1 may appear as the outcome of a number of different types of investigation. Differences in point of view which have been expressed regarding the handling of the figures, concern the probability constructs by aid of which the bare numerical data recorded in the table provide a basis for inference. Two lines of approach may be distinguished.

Table 1

	Col. 1	Col. 2	Total
Row 1	a	c	m
Row 2	b	d	n
Total	r	s	N

Following the first, it is considered that for all the types of problem,* the relevant information on the points at issue may be obtained by comparing the observed pattern of cell contents (a, b, c, d) with the set of possible patterns, all giving the marginal totals actually found in the sample. Thus there is only one degree of freedom among the four cell frequencies, and the relevant probability distribution is obtained from the hypergeometric series. This approach can be derived from Fisher's information theory. But without using this theory, a may be referred to the one-dimensioned, conditional set as a convenient practical device, which avoids the introduction of nuisance parameters.

From the point of view of the second approach it is an over-simplification to treat every case providing data in the form of Table 1 as a problem of sampling with fixed marginal totals. It is suggested that the readiness of the mind to assimilate the information provided by the statistical analysis depends on the directness of the relation between the theoretical probability set and the random process of selection introduced in collecting the data. Since the random procedure may have entered in different ways, the appropriate probability constructs may be expected to differ.

It can be argued that, except in the case of small samples, the difference of approach is practically unimportant, and that even here, until tables of the kind which Barnard has in mind are available, the statistician will be forced to draw his conclusions from a table of the conditional distribution, such as that recently prepared by Finney (1948). But there is another aspect to the matter. The published discussion has hitherto been concerned primarily with the sampling distribution of a statistic under the null hypothesis. If we go beyond this

* Three types were discussed by Barnard (1947) and Pearson (1947).

and consider the sensitivity of the test, that is to say, its power to detect differences if they exist, it is at once clear that all problems cannot be treated in the same manner.

In a recent paper, Patnaik (1948) has considered this aspect of what one of us (Pearson, 1947) has termed Problem II and of what Barnard termed the 2×2 comparative trial. This occurs when we inquire whether the probability of an individual bearing a given character A, is the same in two large populations from which random samples of size m and n, respectively, have been drawn. Here, two separate random selections are involved, and the cell contents may be described as having two degrees of freedom. If the two population probabilities are unequal, i.e. $p_1(A) \neq p_2(A)$, the chance that the test will establish a difference at a given significance level α is a function of p_1 and p_2, of m and n and of α. This relationship was explored by Patnaik.

In the present paper we shall consider what Pearson termed Problem I and Barnard the 2×2 independence trial, from this aspect of the power of the test. In this case, only a single process of random selection or partition is called for.

2. Statement of the Problem

For convenience we shall describe the type of experiment we have in mind as one in which two 'treatments', say A and B, are compared; the response is a quantal one, so that an individual either 'reacts' or 'fails to react'. The applications suggested by these terms lie in the biological field, but there is no difficulty in translating the terms of the theoretical picture to fit a case where, for example, the individual is a shell, the two treatments are two types of fuze and the reaction is successful perforation of a steel plate.

In this Problem I, the N individuals available for experiment are divided by a random partition into a group of m which receive treatment A, and a group of $n = N - m$ which receive treatment B. It is then observed that a/m and b/n react in the specified manner. The experiment is self-contained and the random process under complete control; but without further assumptions or knowledge, the inferences that are possible relate only to the reactions of the N individuals to the treatments. This may be all that is called for. Inferences of wider application may be drawn by assuming that the N have been sampled randomly from a population in which we are interested. Or, as is often the case, the experiment may be one of a related series, each experiment in which is self-contained. These, taken as a whole, can form the basis of reasoned conclusions regarding the treatments, conclusions which are not dependent on all groups of individuals having been drawn randomly from a unique population in the rigorous statistical sense. This is the case if the tests are applied to laboratory animals whose susceptibility may change somewhat from time to time. Or, as Barnard has suggested, when an open-air gunnery trial runs over several days of inconstant weather.

The question then is this: confining attention to the group of N individuals, in what sense is it possible to interpret a difference in treatments and how can we measure the power of the test to detect such a difference? Perhaps the most general method of regarding the problem is to suppose that the N individuals fall into four classes:

(i) those who would react if given either treatment, X in number;
(ii) those who would react only if given treatment A, W in number;
(iii) those who would react only if given treatment B, Y in number;
(iv) those who would react to neither treatment, Z in number.

In this way we recognize that every individual does not respond in the same way to a given treatment; that the success of a shell in perforating a plate will depend not only on the fuzing, but on other factors such as strength of shell-case, angle of yaw on striking, the position of the strike on the plate, etc. These latter factors are purposely randomly associated with the two types of fuze under trial.

If m individuals are selected randomly and assigned treatment A and the remaining n assigned treatment B, the resulting partition will be that shown in Table 2*a*; as, however, we can only observe whether an individual has reacted or not, the figures available for analysis will be in the form of Table 2*b*.

Table 2*a*

	React if given A or B	Only react if given A	Only react if given B	React to neither treatment	Total
Treatment A	x_1	w_1	y_1	z_1	m
Treatment B	x_2	w_2	y_2	z_2	n
Total	X	W	Y	Z	N

Table 2*b*

	React	Fail to react	Total
Treatment A	$x_1+w_1=a$	$y_1+z_1=c$	m
Treatment B	$x_2+y_2=b$	$w_2+z_2=d$	n
Total	$X+w_1+y_2=r$	$Z+w_2+y_1=s$	N

The usual null hypothesis is that the treatments are identical as far as producing a reaction on these N individuals is concerned, i.e. it is the hypothesis that $W = Y = 0$. The test of significance of departure from hypothesis would be applied to the 2×2 Table 2*b*, and in its exact form consists in referring $a = x_1+w_1$ to the appropriate hypergeometric distribution, with parameters r, N and m. Here, if the sample is not too large, Finney's (1948) table is applicable. More approximately, we can either regard

$$\frac{a-rm/N}{\sqrt{\left(\dfrac{mnrs}{N^2(N-1)}\right)}} \quad \text{as a unit normal deviate, or}$$

$$\frac{(ad-bc)^2N}{mnrs} \quad \text{as a } \chi^2 \text{ with 1 degree of freedom,}$$

making a correction for continuity, if necessary.

It should be noted that if treatment A were regarded as more successful than B when $X+W > X+Y$, then the null hypothesis to test would be that $W-Y = 0$. The expectation of the difference $a/m - b/n$ is still zero if $W = Y \neq 0$, but its sampling distribution under random partition can no longer be determined from the marginal totals of Table 2b. The experiment as planned is not, in fact, able to distinguish between the cases $W = Y = 0$ and $W - Y = 0$; to do so would involve applying both treatments to the same individuals, which will usually be impossible in practice.

Table 3a

	React if given A or B	React if given B	React to neither treatment	Total
Treatment A	x_1	y_1	z_1	m
Treatment B	x_2	y_2	z_2	n
Total	X	Y	Z	N

Table 3b

	React	Fail to react	Total
Treatment A	$x_1 = a$	$y_1 + z_1 = c$	m
Treatment B	$x_2 + y_2 = b$	$z_2 = d$	n
Total	$X + y_2 = r$	$Z + y_1 = s$	N

In the discussion which follows we shall suppose that $W = 0$. This may narrow the field of application, but not too seriously. If A is an old treatment and B a new one which it is hoped is an improvement,* we are assuming that B has at least the qualities of A. Thus if B aims at the cure of a disease and A is a control corresponding to 'no treatment', we assume that in no case will B *prevent* a recovery which would have taken place without any treatment at all. With $W = 0$, we have the scheme of Tables 3a and b. In this case the null hypothesis is that $Y = 0$; the alternative that interests us is that $Y > 0$, so that the statistical test applied to Table 3b involves comparing the observed value of a with its lower significance level, or b with its upper level. This critical limit for a may be written $a(\alpha, r, N; m)$, where α is the chance of falling at or below the limit if $Y = 0$. It can be determined either precisely from the hypergeometric or from the normal approximation. If the null hypothesis is not true, $r = X + y_2$ will vary from one random partition to another. Thus the problem is to determine the probability that, when $Y \neq 0$, $a \leqslant a(\alpha, r, N, m)$, where both a and r are random variables.

* For purposes of discussion, we take 'reaction' to be good.

A numerical illustration will help to make the position clear. Thirty small disks were placed in a box, of which $X = 10$ were coloured red, $Y = 10$ coloured green and $Z = 10$ coloured yellow. The disks were divided randomly into two groups A and B each containing $m = n = 15$. In terms of treatment comparisons, reds 'respond to treatment' if they fall into group A, reds and greens if they fall into group B. Four of many possible partitions are shown in the 2×3 tables of row I below; under these are given in row II the corresponding 2×2 tables which contain the numerical data available to the experimenter. He cannot distinguish between greens and yellows in group A or between reds and greens in group B. The null hypothesis is that there were no green disks in the box of 30. Below the 2×2 tables are shown (III) the critical limits corresponding to a nominal 5 % significance level, (IV) the true level or sum of the tail terms and (V) the conclusion that would be drawn. It may be noted that this experimental partition was repeated 50 times and that significance was established, i.e. the presence of green disks inferred, in 20 of these partitions. This is a result which, as will be shown below, agrees closely with expectation.

	Colour ...	R.	G.	Y.		R.	G.	Y.		R.	G.	Y.		R.	G.	Y.	
I	A	3	6	6	15	5	5	5	15	5	4	6	15	7	3	5	15
	B	7	4	4	15	5	5	5	15	5	6	4	15	3	7	5	15
		10	10	10	30	10	10	10	30	10	10	10	30	10	10	10	30
II			3	12	15		5	10	15		5	10	15		7	8	15
			11	4	15		10	5	15		11	4	15		10	5	15
			14	16	30		15	15	30		16	14	30		17	13	30
III	Rejection level, given r	$a_0 = 4$				$a_0 = 4$				$a_0 = 5$				$a_0 = 5$			
IV	$P\{a \leqslant a_0 \mid r\}$	0·0328				0·0134				0·0328				0·0127			
V	Significant	Yes				No				Yes				No			

The probability that $a \leqslant a(\alpha, r, N, m)$ when $Y \neq 0$ represents the power of the test in the sense of Neyman & Pearson, i.e. the probability of establishing significance at the 100α % level when $Y > 0$. It will be a function not only of Y but of X (or Z). Owing to discontinuity in the distribution of a, it is, of course, impossible to choose $a(\alpha, r, N, m)$ with the same value of α for all r. In our calculations given below we have used what may be termed nominal 5 and 15 % levels, such that the critical limits $a(\alpha, r, N, m)$ are the highest integer values satisfying the inequalities

$$P\{a \leqslant a(\alpha, r, N, m)\} \leqslant \alpha \quad (\alpha = 0{\cdot}05 \text{ and } 0{\cdot}15). \tag{1}$$

We have taken $m = n = \frac{1}{2}N$, supposing that each treatment is given to the same number of individuals. This will usually be the case in a planned experiment with randomization, and as our object is to illustrate certain points which we believe are of general interest, no exhaustive tabulation is called for.

The method of calculation leading to Figs. 2–7 is described in the following section and a discussion of results is given in § 4.

3. Method of Calculation

In a 2×3 table of fixed margins, there are two degrees of freedom. We shall take as independent variables x_1 and y_1. The probability of a particular partition of Table 3a is

$$p(x_1, y_1) = \frac{m!\,n!\,X!\,Y!\,Z!}{x_1!\,x_2!\,y_1!\,y_2!\,z_1!\,z_2!\,N!}, \tag{2}$$

where x_2, y_2, z_1 and z_2 can all be expressed in terms of x_1, y_1 and the marginal totals. For mathematical convenience we may regard the partition of Table 3a as obtained in two steps, the first determining y_1, the second x_1. In fact, the expression (2) may be written as a product of two parts

$$\frac{m!\,n!\,Y!(N-Y)!}{y_1!\,y_2!(m-y_1)!\,(n-y_2)!\,N!} \times \frac{X!\,Z!(x_1+z_1)!\,(x_2+z_2)!}{x_1!\,x_2!\,z_1!\,z_2!(X+Z)!}. \tag{3}$$

The first factor, say $p(y_1)$, is the probability of obtaining y_1 individuals with a character A in a sample of Y drawn randomly, without replacement, from a population of N individuals of whom m have a character A. The second factor, say $p(x_1 \mid y_1)$, is the probability of obtaining x_1 with A in a sample of X drawn, without replacement, from the remaining $N-Y$ individuals, of whom $m-y_1$ now have A. Thus

$$p(x_1 y_1) = p(y_1) \times p(x_1 \mid y_1), \tag{4}$$

and while $p(y_1)$ is the marginal distribution for y_1 of the joint distribution $p(x_1 y_1)$, $p(x_1 \mid y_1)$ is the relative probability distribution of x_1 in the array of constant y_1. Both $p(y_1)$ and $p(x_1 \mid y_1)$ are of hypergeometric type; the joint distribution has been discussed by K. Pearson (1924) who termed it a double hypergeometric series and considered how it might lead to a frequency surface.* He gave some of the momental constants of $p(x_1 \mid y_1)$ which we shall use below; in our present notation, and for the special case $m = n = \frac{1}{2}N$, these become

$$\text{Mean } (x_1 \mid y_1) = Xp, \tag{5}$$

$$\text{Variance } (x_1 \mid y_1) = Xpq\frac{Z}{X+Z-1}, \tag{6}$$

$$\beta_1(x_1 \mid y_1) = \frac{(p-q)^2}{Xpq}\frac{\{1-2(X-1)/(X+Z-2)\}^2}{1-(X-1)/(X+Z-1)}, \tag{7}$$

where
$$p = 1-q = (m-y_1)/(N-Y) = (x_1+z_1)/(X+Z). \tag{8}$$

The distribution $p(x_1 y_1)$ is bounded by certain limiting lines, thus:

$$x_1 \leqslant X, \quad y_1 \leqslant Y, \quad m \geqslant x_1 + y_1 \geqslant m - Z. \tag{9}$$

Fig. 1 illustrates the position for the case

$$X = 10, \quad Y = 15, \quad Z = 5, \quad m = n = \tfrac{1}{2}N = 15. \tag{10}$$

In applying the test of significance to the 2×2 Table 3b, we shall reject the null hypothesis ($Y = 0$) when

$$x_1 = a \leqslant a(\alpha, r, N, m). \tag{11}$$

But
$$r = X + y_2 = X + Y - y_1. \tag{12}$$

Hence for given X and Y there will be a critical limit in each y_1 array, such that if x_1 fall below the limit, the hypothesis will be rejected. For the marginal values given in (10),

* He had illustrated the distribution many years before (K. Pearson, 1895) in connexion with the correlation between the number of cards of a given suit held by two players at whist.

Fig. 1 shows these rejection limits as a connected line for the nominal level $\alpha = 0{\cdot}05$.* The power of the test for a given X and Y is then the sum of the probabilities $p(x_1 y_1)$ taken over the points of the field outside (or below) this critical boundary. Fig. 4 shows that for this particular case the sum is about 0·85.

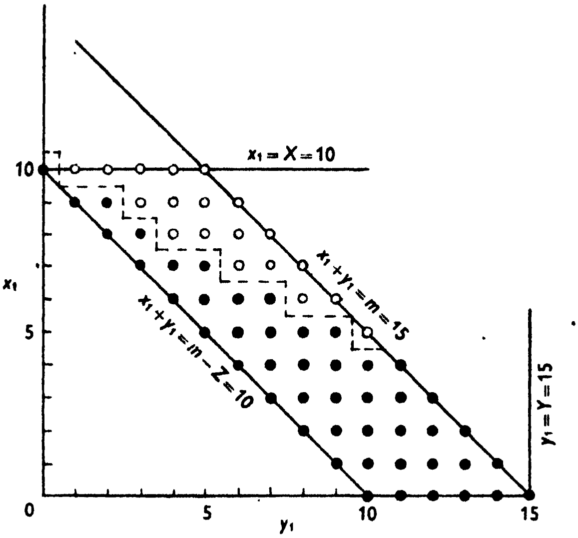


Fig. 1. Illustration for case of partition of 30 individuals into two groups, $m = n = \frac{1}{2}N = 15$; $X = 10$, $Y = 15$, $Z = 5$; nominal significance level, $\alpha = 0{\cdot}05$. The null hypothesis will be rejected at the (x_1, y_1) points marked ●, where $x_1 = a$, $y_1 = 25 - r$.

For small samples the calculations are not very tedious if undertaken systematically, but they become so when N is large. In the case $N = 50$ we have therefore used an approximation based on assuming that the array distributions $p(x_1 \mid y_1)$ can be represented by normal curves with the means and variances of equations (5) and (6). Since, using (8), when $m = \frac{1}{2}N$,

$$q - p = (2y_1 - Y)/(N - Y),$$

where the expectation of y_1 is $\frac{1}{2}Y$, it follows that the β_1 of the array distributions which contain most of the frequency will be small provided that neither X or Z are too small. The accuracy of this normal approximation is illustrated below.

The calculations on which Figs. 2–7 have been based may be summarized as follows:

(i) $N = 20$ *and* 30

Full calculations of the terms of the hypergeometric series

$$\frac{(m!)^2 r!(N-r)!}{a!b!c!d!N!}$$

were made for all values of r, and hence the levels corresponding to the nominal $\alpha = 0{\cdot}05$ and 0·15 were found. This determined the position of the boundary of the critical region as illustrated in Fig. 1 and also provided the terms of the series $p(y_1)$. The expressions $p(x_1 y_1)$ corresponding to points in the critical region were then found and summed for selected values of X, Y and Z, giving the ordinates of the power curves drawn in Figs. 2–5.

* These limits will be found to agree with those in Finney's (1948) table. The relevant section is that for $A = 15 = B$ (in his notation); further, since he takes $a > b$ his (i)a and (ii)b correspond to our (i) $25 - (x_1 + y_1)$ and (ii) x_1. See Appendix, p. 345 below. For example, points in Fig. 1 on the diagonal $x_1 + y_1 = 12$ make Finney's $a = 13$ and for this value he gives the 5 % significance level as $b = 7$; thus all points on this diagonal for which $x_1 \leqslant 7$ fall in the rejection region.

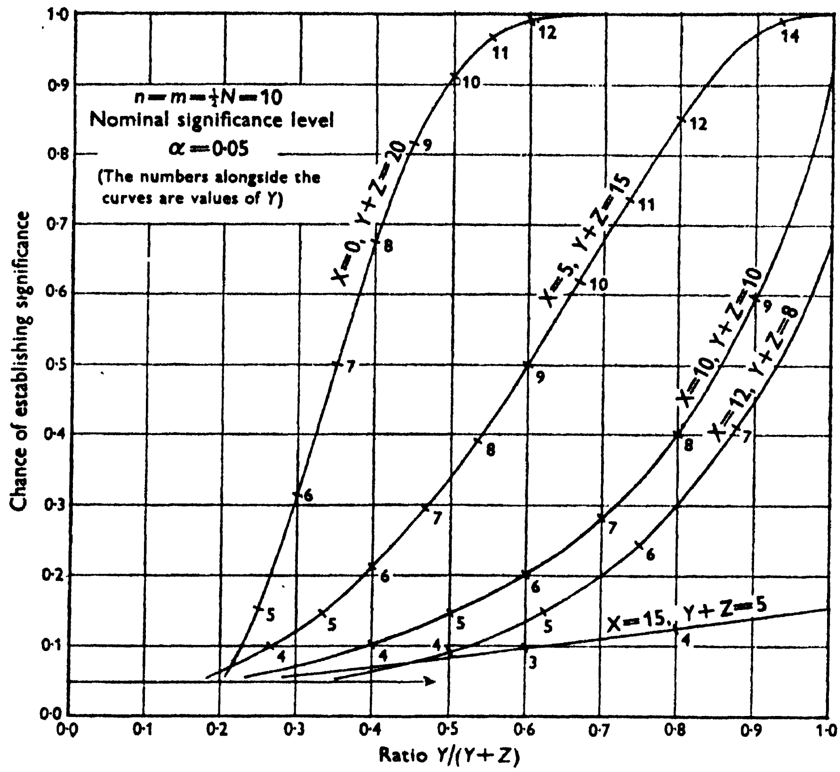

n = m = ½N = 10
Nominal significance level
α = 0·05
(The numbers alongside the curves are values of Y)
X=0, Y+Z=20
X=5, Y+Z=15
X=10, Y+Z=10
X=12, Y+Z=8
X=15, Y+Z=5
Chance of establishing significance
Ratio Y/(Y+Z)

Fig. 2

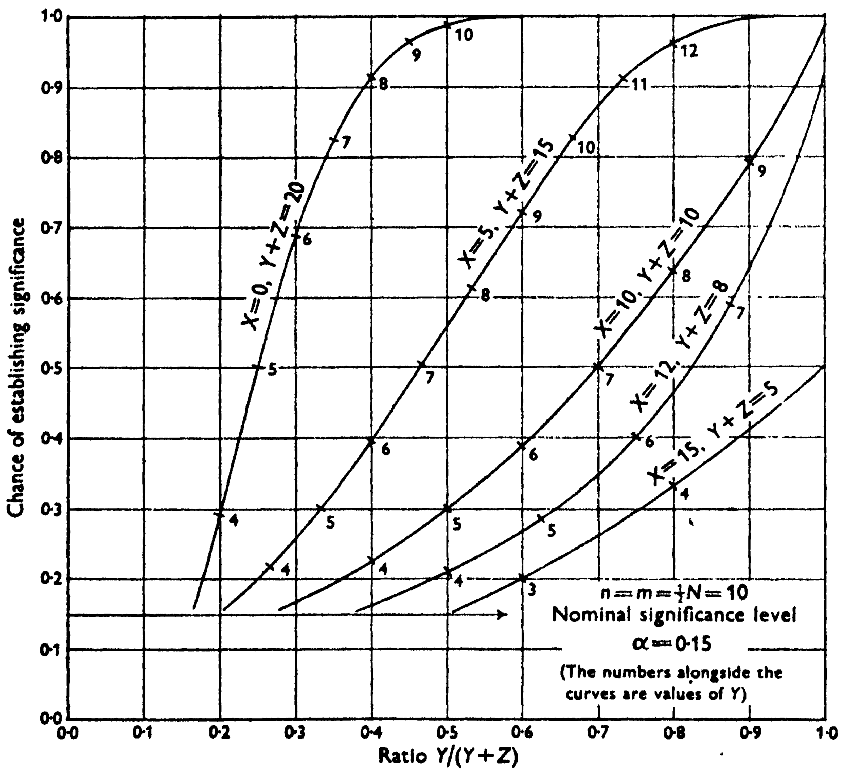

n = m = ½N = 10
Nominal significance level
α = 0·15
(The numbers alongside the curves are values of Y)
X=0, Y+Z=20
X=5, Y+Z=15
X=10, Y+Z=10
X=12, Y+Z=8
X=15, Y+Z=5
Chance of establishing significance
Ratio Y/(Y+Z)

Fig. 3

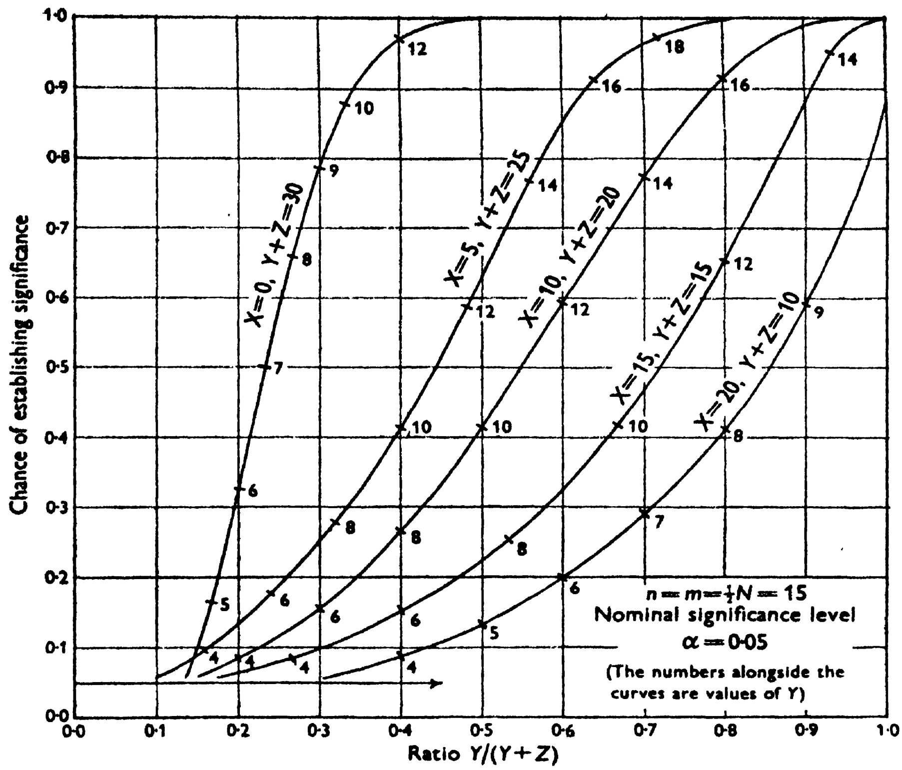

Chance of establishing significance
Ratio Y/(Y+Z)
X=0, Y+Z=30
X=5, Y+Z=25
X=10, Y+Z=20
X=15, Y+Z=15
X=20, Y+Z=10
n = m = ½N = 15
Nominal significance level
α = 0·05
(The numbers alongside the curves are values of Y)

Fig. 4

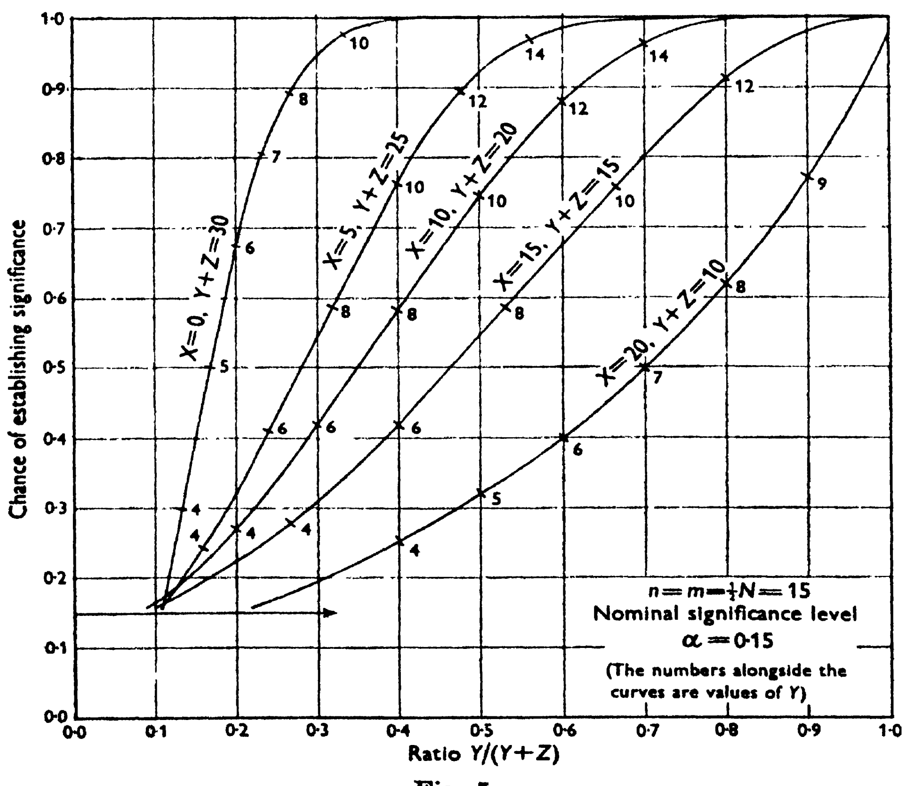

Chance of establishing significance
Ratio Y/(Y+Z)
X=0, Y+Z=30
X=5, Y+Z=25
X=10, Y+Z=20
X=15, Y+Z=15
X=20, Y+Z=10
n = m = ½N = 15
Nominal significance level
α = 0·15
(The numbers alongside the curves are values of Y)

Fig. 5

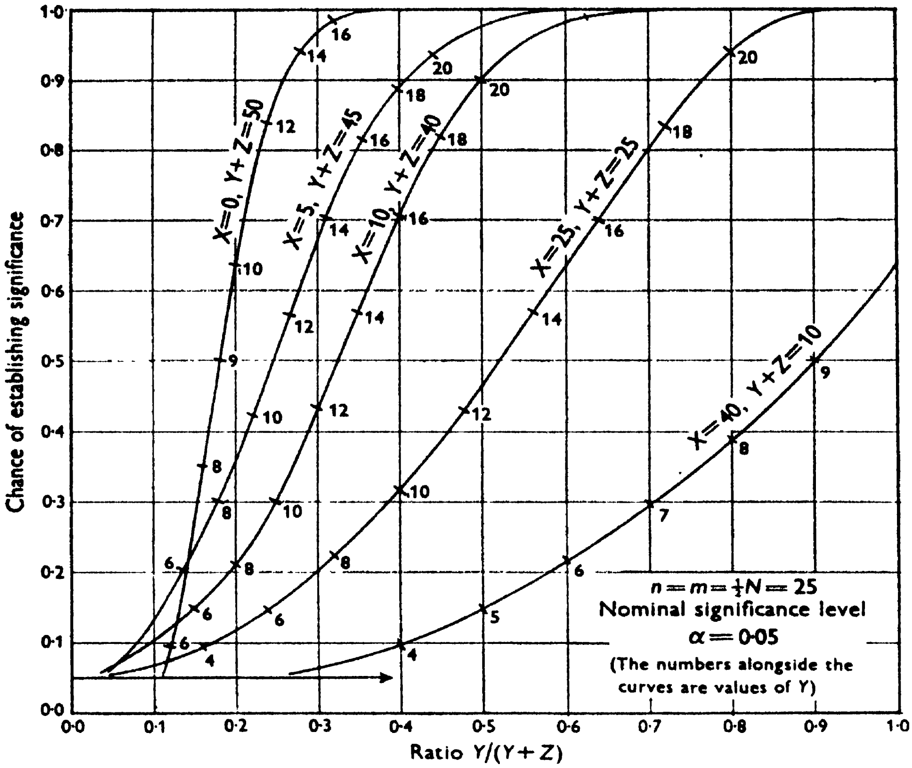

Chance of establishing significance
Ratio Y/(Y+Z)
X=0, Y+Z=50
X=5, Y+Z=45
X=10, Y+Z=40
X=25, Y+Z=25
X=40, Y+Z=10
n=m=½N=25
Nominal significance level
α=0·05
(The numbers alongside the curves are values of Y)

Fig. 6

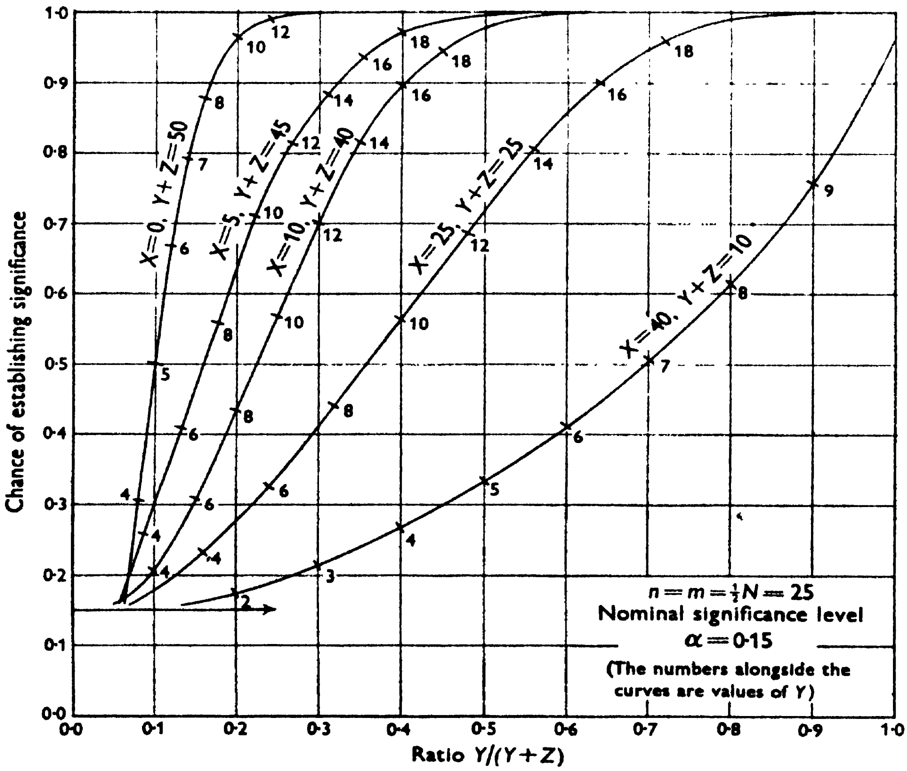

Chance of establishing significance
Ratio Y/(Y+Z)
X=0, Y+Z=50
X=5, Y+Z=45
X=10, Y+Z=40
X=25, Y+Z=25
X=40, Y+Z=10
n=m=½N=25
Nominal significance level
α=0·15
(The numbers alongside the curves are values of Y)

Fig. 7

(ii) *Check for* $N = 30$

For a number of cases the normal approximation referred to above was also used. This involved finding for each array the integral under the normal curve with mean and variance, (5) and (6), beyond the point $a(\alpha, r, N, m) + 0{\cdot}5$, using a correction for continuity. These integrals were multiplied by the corresponding exact value of $p(y_1)$ and summed for all arrays. A comparison of exact and approximate results is shown in Table 4. From our present point of view the agreement may be regarded as good.

Table 4. *Approximate solution for power function. Case* $m = n = \frac{1}{2}N = 15$

X	Y	$\alpha = 0{\cdot}05$		$\alpha = 0{\cdot}15$	
		True power	Approx.	True power	Approx.
5	2	0·047	0·053	0·121	0·126
	5	·135	·139	·321	·324
	10	·413	·414	·764	·758
	15	·852	·844	·984	·979
	20	·999	·997	1·000	·999
10	2	0·039	0·043	0·166	0·170
	6	·155	·161	·420	·421
	10	·413	·415	·745	·740
	14	·773	·767	·962	·956
	18	·989	·989	1·000	1·000
15	1	0·033	0·036	0·136	0·140
	3	·064	·067	·224	·227
	6	·152	·157	·419	·420
	9	·326	·329	·676	·673
	12	·655	·651	·913	·908
	14	·951	·948	·990	·989

(iii) $N = 50$

Encouraged by the check for $N = 30$, the approximate method was employed throughout in this case, a few checks only being made by the full method, which now becomes rather laborious. The resulting power curves are shown in Figs. 6 and 7.

As N increases above 50, it is probable that further simplifying approximations could be introduced, but these we have not explored. It would also be possible to form diagrams giving contours of constant power, similar to those provided by Patnaik (1948) in his Figs. 2 and 3, for the case where samples are drawn independently from two populations. The parameters corresponding to his p_1 and p_2 would now seem to be X/N and $(X+Y)/N$, the proportion of individuals among the N who would respond to treatments A and B, respectively. A preliminary examination on the basis of the results used to form Figs. 2–7, suggests that the contours corresponding to various values of N, α and the power, may all belong approximately to a single family of curves, as was the case in Patnaik's problem, for $m = n$.

4. Discussion

In the first place some comment is necessary on the significance levels chosen. Owing to the discontinuity which occurs because the cell contents can assume integer values only, the standard levels of 10, 5, 1 %, etc., have not their ordinary meaning. As an example, take the case where $m = n = \frac{1}{2}N = 25$, $r = 16$. On the null hypothesis we have:

$$\begin{aligned} \text{Chance that } a \leqslant 6 &\quad \text{is} \quad 0{\cdot}182, \\ \leqslant 5 &\quad \text{is} \quad 0{\cdot}064, \\ \leqslant 4 &\quad \text{is} \quad 0{\cdot}016, \\ \leqslant 3 &\quad \text{is} \quad 0{\cdot}003. \end{aligned}$$

Thus if we wish to accept *no more than* a 1 in 20 risk of rejecting the hypothesis when it is true, we should only do so when $a \leqslant 4$. But in practice, if prepared to accept a risk of *about* 1 in 20, we should clearly take $a = 5$ as the limit. Similarly, we should take $a = 4$ for a risk of *about* 1 in 100. Thus we should tend to base our conclusions on the actual tail sum found after the data are collected. But if we use the power function concept to inquire *in advance* how large N must be to make an experiment worth while, then we must think in terms of a specific upper limit or nominal level α. Hence it is important to know how much on the average the true level falls short of the nominal. Table 5 has been prepared to indicate this difference in the cases with $N = 20$, 30 and 50 and for nominal levels of $\alpha = 0{\cdot}05$ and $0{\cdot}15$. Finney's (1948) table also brings this out.*

These upper limit values are well on the safe side, and this may be what is wanted if we attach prime importance to Neyman & Pearson's first kind of error; i.e. to the risk of assuming a difference when none exists. But where it is of first importance to find a new solution, e.g. an improvement in treatment, higher risks in this direction will be accepted in order to avoid the chance of falling into the second kind of error, i.e. of overlooking a real difference when it exists. For this reason our calculations have been made for $\alpha = 0{\cdot}05$ and also for the rather high value $\alpha = 0{\cdot}15$, which, as can be found from Table 5, means on the average† a true level of $\bar{\alpha} = 0{\cdot}057$ for $N = 20$, $\bar{\alpha} = 0{\cdot}082$ for $N = 30$, $\bar{\alpha} = 0{\cdot}089$ for $N = 50$.

Turning to the interpretation of the diagrams, it must be emphasized again that the frequencies X, Y and Z will of course not be known; the purpose of the charts is to show for certain sample sizes, the combinations of these three frequencies needed to give a reasonable chance of establishing a significant treatment difference. If the investigation has the positive objective of establishing the value of new methods, it will naturally be hoped that the comparative experiment will establish statistical significance. If, with the knowledge available, it can be shown that an experiment of given magnitude is very unlikely to lead to a significant result, then it may be a waste of time to proceed on this scale. We think that the results presented in the diagrams will be helpful in this type of review.

If treatment A has already been in use, some information will be available as to the likely value of X/N, the proportion of individuals in a group of N who would react to A. We shall at least know whether it is more likely to be 0 or $\frac{1}{2}$. The ratio $Y/(Y+Z)$ used as the abscissa for the power curves measures the headway which the new treatment B could make in causing a satisfactory response among individuals with whom treatment A fails. Again, the experimenter will generally have some idea of what he hopes the treatment will achieve.

* See Appendix, p. 345 below.

† Giving equal weight to all values of r.

He may know, for example, that $Y/(Y+Z)$ is unlikely to exceed 0·25, and yet be clear that even if it were no larger than this the introduction of B would be amply justified. Here Figs. 4 and 5 show that, for $m = n = 15$, were $X = 0$ (treatment A ineffective*) and $Y/(Y+Z) = 8/30 = 0{\cdot}27$, the chance of establishing significance is about 0·66 at the nominal 5 % level and 0·89 at the nominal 0·15 % level. But if X were 10, or one-third of the group of $N = 30$ were responsive to the old treatment then, though 5 of the remaining 20 would

Table 5. *Showing for the case* $m = n = \frac{1}{2}N$: (i) *the critical limits* $a_0(\alpha, r, N, m)$ *for nominal significance levels* $\alpha = 0{\cdot}05$ *and* $0{\cdot}15$; (ii) *the true chance that* $a \leqslant a_0(\alpha, r, N, m)$

r	$N = 20$				$N = 30$				$N = 50$			
	$\alpha = 0{\cdot}05$		$\alpha = 0{\cdot}15$		$\alpha = 0{\cdot}05$		$\alpha = 0{\cdot}15$		$\alpha = 0{\cdot}05$		$\alpha = 0{\cdot}15$	
	a_0	True chance	a_0	True chance	a_0	True chance	a_0	True chance	a_0	True chance	a_0	True chance
3	—	—	0	0·105	—	—	0	0·112	—	—	0	0·117
4	0	0·043	0	0·043	0	0·050	0	0·050	—	—	0	0·055
5	0	0·016	0	0·016	0	0·021	0	0·021	0	0·025	0	0·025
6	0	0·005	1	0·070	0	0·008	1	0·084	0	0·011	1	0·095
7	1	0·029	1	0·029	1	0·040	1	0·040	1	0·049	1	0·049
8	1	0·010	2	0·085	1	0·018	2	0·107	1	0·024	2	0·123
9	2	0·035	2	0·035	1	0·007	2	0·054	1	0·012	2	0·069
10	2	0·012	3	0·089	2	0·025	3	0·123	2	0·037	3	0·144
11	—	—	—	—	2	0·010	3	0·064	2	0·019	3	0·085
12	—	—	—	—	3	0·030	4	0·132	3	0·048	3	0·048
13	—	—	—	—	3	0·013	4	0·070	3	0·025	4	0·098
14	—	—	—	—	4	0·033	5	0·136	3	0·013	4	0·057
15	—	—	—	—	4	0·013	5	0·071	4	0·031	5	0·108
16	—	—	—	—	—	—	—	—	4	0·016	5	0·064
17	—	—	—	—	—	—	—	—	5	0·038	6	0·119
18	—	—	—	—	—	—	—	—	5	0·021	6	0·072
19	—	—	—	—	—	—	—	—	6	0·042	7	0·124
20	—	—	—	—	—	—	—	—	6	0·023	7	0·077
21	—	—	—	—	—	—	—	—	7	0·044	8	0·128
22	—	—	—	—	—	—	—	—	7	0·024	8	0·079
23	—	—	—	—	—	—	—	—	8	0·046	9	0·131
24	—	—	—	—	—	—	—	—	8	0·025	9	0·081
25	—	—	—	—	—	—	—	—	9	0·046	10	0·131

For $r > \frac{1}{2}N$ the true chance for r is the same as that for $N-r$, while $a_0(\alpha, r, N, m) = r - m + a_0(\alpha, N-r, N, m)$.

react if given treatment B ($Y/(Y+Z) = 0{\cdot}25$), the experiment will most probably be inconclusive. This is because the power of the test is now only 0·12 and 0·34, respectively, at the 5 and 15 % levels.

If we regard $Y/(Y+Z)$ as the measure of effectiveness, the existence of X reduces the chance of establishing significance. We need not express effectiveness in this way, but the numerical data given in the charts associating power with the partition of N in X, Y and Z, makes it possible to express the position in any terms considered more appropriate.

As a further example, we may take the case where out of $N = 30$ individuals, $X = 10$ would react to treatment A and $X + Y = 20$ to treatment B. This might well be regarded

* At least, ineffective on the 30 individuals selected for the experiment.

as an eminently satisfactory result. But Figs. 4 and 5 show that the chance of distinguishing this case from one in which $Y = 0$ is only 0·41 at the 5 % level and 0·74 at the 15 % level.* If now we make a rough interpolation in Figs. 6 and 7 between the power curves for $X = 10$ and $X = 25$, it is seen that using 50 individuals for the experiment, then were $X = Y = 17$ (keeping the same proportions as before) there would be a chance of about 0·76 of establishing significance at the 5 % level and of about 0·92 at the 15 % level. There are likely to be circumstances where, with this knowledge, it would rightly be concluded that while a comparative experiment on two sets of 15 individuals would not be worth undertaking, an experiment using two sets of 25 would be.

Another point which the diagrams bring out is the dilemma which faces the medical research worker who wishes to establish his conclusions on a scientific basis, but has a strong conviction that a new treatment will be effective in reducing pain, hastening recovery or even saving life. Dealing with a hospital population of varying susceptibilities and with other changes in external conditions from time to time, it may be impossible to make comparisons which could be accepted indisputably between (*a*) successes in the past, using treatment A, and (*b*) successes at the moment using treatment B. Yet if a controlled, comparative test is carried out, it is seen that to provide a conclusive result the number of patients in the group Y must be considerable, perhaps 10, 16 or even 20. And yet, on the average, $\frac{1}{2}Y$ of these patients will be assigned the old treatment. Even if the belief in treatment B is based on intuition rather than evidence, considerations of ethics are likely to outweigh the urge to plan an experiment which makes a valid scientific comparison possible.

In conclusion, we freely admit that the presentation given in Figs. 2–7 cannot be regarded as a final one. It does, however, provide enough material for the statistician to consider whether the approach is practically useful and, hence, whether its extension and possible simplification are desirable.

REFERENCES

Barnard, G. A. (1947). *Biometrika*, **34**, 123.
Finney, D. J. (1948). *Biometrika*, **35**, 145.
Patnaik, P. B. (1948). *Biometrika*, **35**, 157.
Pearson, E. S. (1947). *Biometrika*, **34**, 139.
Pearson, K. (1895). *Philos. Trans.* A, **186**, 411.
Pearson, K. (1924). *Biometrika*, **16**, 172.

* In the experiment with disks referred to on p. 335 above, in 20 partitions out of 50 significance was established at the 5 % level and in 40 out of 50 at the 15 % level, results clearly consistent with these theoretical chances.

APPENDIX

Note on the arrangement of D. J. Finney's table (*Biometrika*, **35**, 145–56)

In the present paper we have been mainly concerned with a lower significance level for a, for a given value of $r = a+b$. This limiting value of a is determined by the sum of the tail terms in a hypergeometric series. Thus in Fig. 8 below, for $m = n = \frac{1}{2}N = 15$, when $r = 13$, what we have termed the nominal lower 5 % significance level for a is 3 and when $r = 12$ it is also 3. This is because the sum of the chances of a assuming values of 0, 1, 2 and 3 on the null hypothesis is 0·0127 in the former case and 0·0301 in the latter, while it would be over 0·05 if we included the chance that $a = 4$. On this basis, the stepped line indicates the position of the significance level for different values of r.

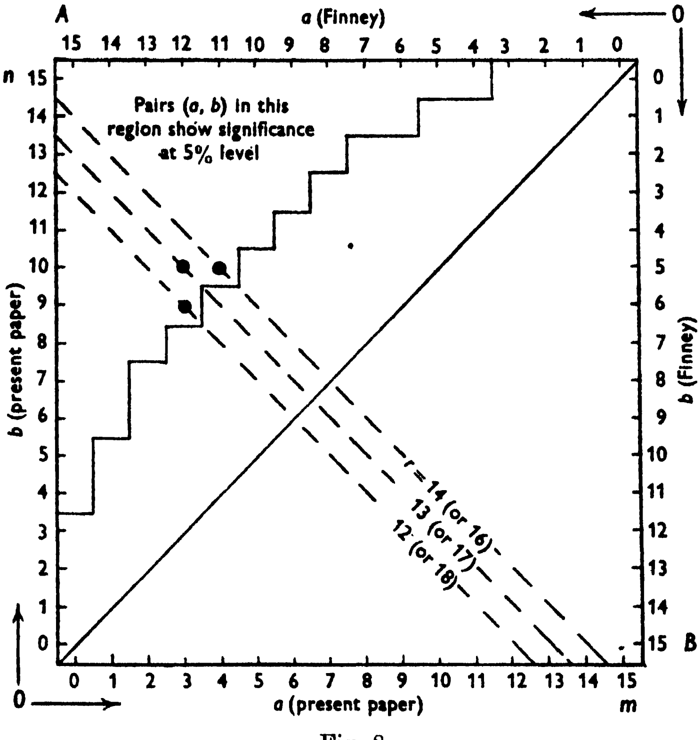


Fig. 8

In Finney's arrangement, when dealing with the single-tail test, $a/A \geqslant b/B$, where $A = m$ and $B = n$. Thus in Fig. 8 the scales of a and b must be reversed. His table is not entered with r, but shows for a given a the highest value of b which is just significant at the 5 % level. Thus when $A = B = 15$, for $a = 11$ he gives $b = 5$, and for $a = 12$ he gives $b = 6$. These points correspond to $r = 16$, $a = 11$ and $r = 18$, $a = 12$ in the diagram. It will be seen that on this basis an entry for the point $a = 12$, $b = 5$ on the intermediate diagonal with $r = 17$ is unnecessary and space is saved in tabulation. But there is no difference whatsoever in the basic calculations leading to Finney's table and to the special results we have given in Table 5 above.

STATISTICAL ANALYSIS OF A NON-ORTHOGONAL TRI-FACTORIAL EXPERIMENT

BY W. L. STEVENS, *Universidade de São Paulo, Brazil*

1. INTRODUCTION

The formal mathematical solution to the problem of 'fitting constants' to a set of non-orthogonal data presents no great difficulty, but at first sight the arithmetical labour of solving the normal equations may appear so great as to discourage the experimenter. It is therefore important to show, by suitable worked examples, that the arithmetical procedure can be reduced to a routine so that, although the computations will always be lengthy, they can at least be seen to follow a simple and recognizable pattern.

The classical treatment of the problem by the 'method of least squares' pays scant attention to the formulation of valid tests of significance. When interactions of various factors have to be tested for significance, the mathematical solution is likely to appear very complex and much less easy to understand than the arithmetic of a worked example.

Special cases of non-orthogonality, such as those arising from a 'missing plot' or a missing row in a Latin square, have been treated by Fisher, Yates and other writers who have produced elegant solutions, but we are concerned here with a breakdown of the orthogonality restrictions so complete that no special devices are of service. The example to be discussed arose from an experiment made by Dr Rocha Faria of Lisbon, and the data were kindly placed at the disposal of the writer by the Portuguese statistician, Sr Augusto J. de Oliveira. Since it would be difficult even to invent an example more suitable for an illustration of the methods, opportunity has been taken to demonstrate the full arithmetical procedure necessary to effect an analysis of these data.

We may begin by reminding the reader of what is meant by 'fitting constants'. Suppose, for example, that only two factors are being studied—sex and diet, i.e. we have a number of animals of the two sexes and we divide them into three groups to be fed on three diets, A, B and C respectively, and record the gains in weight during a specified period. If the two sexes, even on the same diet, grow at different rates, and if in addition the diets produce different effects, it is possible that the *expected* gain in weight will be given by the formula

$$E + E_S + E_D,$$

where E is a constant, E_S is one of two constants, according to whether the animal is male or female, and E_D is one of three constants according to which of the three diets is given. The observed gain in weight will, of course, differ from the expected gain, owing to unassignable biological variation, errors of weighing, etc., but we are supposing that this 'residual' variation is random and would average zero over a sufficiently large number of animals.

The constants appearing in the above formula may be termed 'effects', since E_S measure the effect of sex and E_D measure the effects of diet. The first constant E may be called, for want of a better name, the 'general effect'. It will be noticed that they are to some extent arbitrary, for if, say, we increase every E_D by 5 units and diminish E by 5 units, the answers given by the formula will be unaltered. In fact, it is only the differences between the two values of E_S and between the three values of E_D which will be measured by an experiment.

The ambiguity could, of course, be removed by saying that the sum of the two values of E_S or of the three of E_D shall be zero, but generally we see no advantage in adopting such a convention.

It will also be noticed that the adoption of such a formula implies that the effects of sex and diet are additive, i.e. that the sex difference is the same on all diets or (equivalently) the dietary differences are the same in both sexes. Such an assumption may or may not be justified—it is possible, for example, that males respond best to one diet while females respond best to another. In general, if the effects are not additive, the formula required would be

$$E+E_{S,\,D},$$

where $E_{S,\,D}$ is one of six constants corresponding to the six combinations of two sexes and three diets.

If the first formula is adequate, the two factors, sex and diet, are said to be *independent*; if the second formula is needed, there is said to be an *interaction* between sex and diet. It is seen that the analysis of the data poses two problems:

(*a*) What type of formula is needed?

(*b*) What are the best estimates of the numerical values of the constants in this formula?

The solution of both problems is very easy when the design of the experiment ensures orthogonality between sex and diet. For orthogonality, it is not necessary that the numbers in all six classes (generated by two sexes × three diets) should be equal; it is sufficient if they obey a simple rule of proportionality illustrated in the following example:

Numbers of test animals

Diet	A	B	C	Total
Male	8 (= 2 × 4)	10 (= 2 × 5)	6 (= 2 × 3)	24 (= 2 × 12)
Female	12 (= 3 × 4)	15 (= 3 × 5)	9 (= 3 × 3)	36 (= 3 × 12)
Total	20 (= 5 × 4)	25 (= 5 × 5)	15 (= 5 × 3)	60 (= 5 × 12)

When, however, the orthogonality relation no longer holds, there is no simple method of testing the significance of interaction and of estimating the required constants. It becomes necessary to solve large sets of simultaneous linear equations which is only practicable if a routine can be developed which will converge on the solutions through a series of approximations. Furthermore, a valid analysis of variance can no longer be constructed by the standard methods with which the reader is probably familiar. When, as in the example to be discussed, there are more than two factors in the experiment, these difficulties are aggravated.

2. The experiment and data

The experiment was designed to observe the effect on the growth of guinea-pigs of four diets distinguished by the four types of wheat which they included:

A = soft (70 %), B = soft (100 %), C = hard (70 %), D = hard (100 %).

The guinea-pigs were of two sexes and drawn from four litters, so there are three factors: sex, diet and litter.

The data are shown in Table 1, where the complete lack of orthogonality will be apparent. Although no particular interest attaches to the difference between litters, yet the effect of litter will have to be measured in order that it may be eliminated from comparisons between sexes and comparisons between diets. Thus for the purpose of analysis the experiment is tri-factorial.

It is pertinent to describe how, with the material at his disposal, the experimenter might have designed an experiment free from the disadvantages which usually attach to non-orthognality. From the guinea-pigs used in the experiment, it is possible to make up three female and seven male *triads*, i.e. sets of three of the same sex and from the same litter. A further litter would probably have provided a fourth female and an eighth male triad. From the four diets one can construct four sets of three by dropping one at a time. The sets of three diets are then allotted to the triads of animals according to the following scheme:

Litter	Males	Females
I	ABC ABC	ABC
II	BCD BCD	BCD
III and V	CDA CDA	CDA
IV	DAB DAB	DAB

Within each triad the designated three diets are allocated at random.

The experiment described (an example of *balanced incomplete blocks*) is not orthogonal but the departures from orthogonality follow a certain regular pattern, as a consequence of which the statistical analysis, while not so simple as for an orthogonal design, are nevertheless relatively straightforward and follow a standard method.

Other 'balanced' designs are available; the one illustrated was chosen because it gives information of high precision on the effects of diets, of sexes and of the interaction between diet and sex, while yielding only relatively poor information on the differences between litters.

Balanced designs have been described in order to emphasize that modern resources are quite adequate for dealing with the kind of difficulty which confronted this experimenter, and that unbalanced non-orthogonality with its attendant complications could easily have been avoided if he had consulted a statistician *before* instead of *after* doing the experiment. However, this is no answer to a research worker who has spent time and money on his experiment and has discovered too late that better methods are available. The statistician must provide a technique for analysing the data and hope that the arithmetical labours required will be sufficient to discourage the experimenter from ever again disregarding the principles of good experimental design.

3. Preliminary analyses of variance

The first stage in the computation, set out in Table 1, is almost self-explanatory. There is some replication present in the experiment, i.e. animals which belong to the same litter and sex and receive the same diet, and as the replicates are all pairs, the 'error sum of squares' may be found from the sum of squares of the differences divided by two. The sum of squares

Table 1. *Gains in weight (g.) of thirty-six guinea-pigs and first stage of the computation*

Type of wheat in diet	I M.	I F.	II M.	II F.	III M.	III F.	IV M.	IV F.	Total M.	Total F.	Total Both
A = soft (70%)	43	58	73	—	81	62	67	71	381	191	572
	58		59						(6)	(3)	(9)
B = soft (100%)	93	60	75	71	101	76	100	—	541	207	748
	83		89						(6)	(3)	(9)
C = hard (70%)	91	70	85	70	92	—	106	73	462	271	733
				58	88				(5)	(4)	(9)
D = hard (100%)	89	—	98	69	105	—	109	76	598	217	815
	89			72	108				(6)	(3)	(9)
Total	546	188	479	340	575	138	382	220	1982	886	
	(7)	(3)	(6)	(5)	(6)	(2)	(4)	(3)	(23)	(13)	
	734 (10)		819 (11)		713 (8)		602 (7)				$X = 2868$ (36)

Note. The numbers in brackets are the numbers of observations contributing to the respective totals.

Grand total,	$X =$ 2,868
Number of observations,	$n =$ 36
Sum of squares of the observations,	$S(x^2) =$ 238,204
	$X^2/n =$ 228,484
Sum of squares of deviations,	$S(x-\bar{x})^2 =$ 9,720

Differences within the nine pairs which are alike in all respects:

	I M.	I F.	II M.	II F.	III M.	III F.	IV M.	IV F.
A	15	—	14	—	—	—	—	—
B	10	—	14	—	—	—	—	—
C	—	—	—	12	4	—	—	—
D	0	—	—	3	3	—	—	—

Sum of squares of differences, $15^2 + 10^2 + \ldots + 3^2 = 895$

Hence entry for analysis of variance table $= 895/2 = 448$

Preliminary analysis of variance

	Degrees of freedom	Sum of squares	Mean square	Significance
Between 'combinations'	26	9272	357	***
Error (i.e. within pairs)	9	448	49·8	
Total	35	9720		

Conclusion. There is decisive evidence of differences between the twenty-seven represented combinations.

Table 2. *Condensations of the data to two-factor tables, and computation of sums of squares*

Diet	Litter I	II	III	IV	Total
A	159 (3)	132 (2)	143 (2)	138 (2)	572 (9)
B	236 (3)	235 (3)	177 (2)	100 (1)	748 (9)
C	161 (2)	213 (3)	180 (2)	179 (2)	733 (9)
D	178 (2)	239 (3)	213 (2)	185 (2)	815 (9)
Total	734 (10)	819 (11)	713 (8)	602 (7)	2868 (36)

Diet	Sex Male	Female	Total
A	381 (6)	191 (3)	572 (9)
B	541 (6)	207 (3)	748 (9)
C	462 (5)	271 (4)	733 (9)
D	598 (6)	217 (3)	815 (9)
Total	1982 (23)	886 (13)	2868 (36)

Sex	Litter I	II	III	IV	Total
Male	546 (7)	479 (6)	575 (6)	382 (4)	1982 (23)
Female	188 (3)	340 (5)	138 (2)	220 (3)	886 (13)
Total	734 (10)	819 (11)	713 (8)	602 (7)	2868 (36)

Note. Numbers in brackets are the numbers of observations contributing to the respective subtotals.

Sums of squares:

Single factors: Sexes	Diets	Litters	Pairs of factors: Diets, Litters	Litters, Sexes	Sexes, Diets
231,181	232,022	230,172	234,507	232,970	235,763
228,484	228,484	228,484	228,484	228,484	228,484
2,697	3,538	1,688	6,023	4,486	7,279

Sums of squares for interactions:

Diets	3538	Litters	1688	Sexes	2697
Litters	1688	Sexes	2697	Diets	3538
Interaction (by difference)	797	Interaction	101	Interaction	1044
Diets, litters	6023	Litters, sexes	4486	Sexes, diets	7279

Table 3. *Second analysis of variance (approximate)*

Source of variation	Degrees of freedom		Sum of squares		Mean square	Variance ratio	Significance
Sex	1		2697		2697	54·2	***
Diet	3		3538		1179	23·7	***
Litter	3		1688		563	11·3	**
		7		7923	1132		
2-factor interactions:							
Diet/litter	9		797		88·6	1·78	*n.s.*
Litter/sex	3		101		37·0	0·74	*n.s.*
Sex/diet	3		1044		348	6·99	**
		15		1942			
3-factor interaction: sex/diet/litter		4		−593	(?)		
Between combinations		26		9272			
Error		9		448	49·8		
Total		35		9720			

between combinations of litter, sex and diet is then found by subtraction, thus reversing the usual procedure. Here and elsewhere the following code is used for denoting significance:

***	virtual certainty	$P < 0{\cdot}1\,\%$
**	highly significant	$0{\cdot}1\,\% < P < 1\,\%$
*	significant	$1\,\% < P < 5\,\%$
[?]	suggestive	$5\,\% < P < 20\,\%$
n.s.	not significant	$20\,\% < P$

The short analysis of variance shows that some at least of litter, sex and diet must be exerting significant effects and we can proceed to examine these in more detail.

The analysis of variance into the items corresponding to the three primary factors and their interactions is a simple extension of the method used for an ordinary orthogonal three-factor experiment. The data is condensed (Table 2) into three two-way tables showing the subtotals for all combinations of diet × litter, of sex × diet and of litter × sex. Since, as a consequence of non-orthogonality, the number of observations contributing to these subtotals will vary, the actual numbers should be inserted in brackets below the corresponding subtotals (see Table 2).

The sum of squares for any primary factor is found in the way usual when the numbers of observations are unequal, e.g. for 'between litters' we have

$$\frac{734^2}{10}+\frac{819^2}{11}+\frac{713^2}{8}+\frac{602^2}{7}-\frac{2868^2}{36}=1688.$$

Corresponding to any pair of factors, we may find a sum of squares between all combinations of the two factors. Thus between combinations of diet and litter the sum of squares is

$$\frac{159^2}{3}+\frac{132^2}{2}+\frac{143^2}{2}+\ldots+\frac{185^2}{2}-\frac{2868^2}{36}=6023.$$

From this sum we may remove the sums of squares corresponding respectively to the two single factors, diet and litter, and leave something which may be labelled 'interaction between diet and litter'.

Notice however that, as a consequence of non-orthogonality, the primary factors are not wholly independent, and the items labelled 'interaction' are therefore not true interactions nor even true 'sums of squares'. This becomes very evident when we calculate the three-factor interaction by difference (see Table 3) and find that we have run into serious trouble in respect of both degrees of freedom and 'sum of squares'. Only 4 degrees of freedom remain (although the three-factor interaction should have $3 \times 3 \times 1 = 9$ degrees of freedom) and the 'sum of squares' appears negative!

It is evident that the analysis of variance displayed in Table 3 is not genuine, but it will nevertheless prove a useful guide to decide what kind of formula will be needed to represent the effects of the various factors. If, therefore, we take the analysis of variance at its face value, it would appear that all three primary factors have significant effects, and that it is likely also that a significant interaction will be found between sex and diet.

We shall therefore begin by finding sets of 'constants' appropriate to three independent primary factors and afterwards examine the interaction between sex and diet.

4. Effects of three independent factors

If the three factors are independent, the expected value of any observation is of the form

$$E + E_S + E_D + E_L,$$

where E is a constant, E_S is one of two constants for the two sexes, E_D is one of four constants for the four diets, and E_L is one of four constants for the four litters. We do not assume that the constants within any set add to zero, and for practical purposes it is quite immaterial if, for example, every E_S is increased by 5 units and every E_D diminished correspondingly by 5 units.

The computation of the 'effects' (as we have termed these constants) is set out in Table 4. The first step (*i*) is to note, as a guide in subsequent computations, the breakdown of the numbers of observations contributing to the various marginal subtotals. Thus under 'Diets' and against A we find:

$$9 = 3+2+2+2 = 6+3,$$

meaning that there were 9 guinea-pigs on diet A and that these were drawn 3, 2, 2 and 2 from litters I, II, III and IV respectively, and that they consist of 6 males and 3 females.

The marginal subtotals are drawn from Table 2 and set out in (*ii*) of Table 4. The computational procedure may be regarded as answering the question: 'What must I add to or deduct from each observation, in respect of sex (diet, litter) in order to ensure that the means of both sexes (all diets, all litters) shall be equal?'

Thus the means for the two sexes are

$$\begin{array}{ll} \text{Males} & 1982 \div 23 = 86{\cdot}17 \\ \text{Females} & 886 \div 13 = 68{\cdot}15 \\ & \qquad\qquad\quad 18{\cdot}02 \end{array}$$

The average male exceeds the average female by 18·02 g. Two significant figures are sufficient at this stage, so we shall accordingly *deduct* 18 from every *male* observation. The individual adjusted observations are not required, and it is only necessary (and easier) to

Table 4. *Calculation of adjustments for three independent factors (first cycle)*

i Composition of the subtotals:

Sexes		Diets		Litters	
	(diets) (litters)		(litters) (sexes)		(sexes) (diets)
M.	23 = 6 + 6 + 5 + 6 = 7 + 6 + 6 + 4	A	9 = 3 + 2 + 2 + 2 = 6 + 3	I	10 = 7 + 3 = 3 + 3 + 2 + 2
F.	13 = 3 + 3 + 4 + 3 = 3 + 5 + 2 + 3	B	9 = 3 + 3 + 2 + 1 = 6 + 3	II	11 = 6 + 5 = 2 + 3 + 3 + 3
		C	9 = 2 + 3 + 2 + 2 = 5 + 4	III	8 = 6 + 2 = 2 + 2 + 2 + 2
		D	9 = 2 + 3 + 2 + 2 = 6 + 3	IV	7 = 4 + 3 = 2 + 1 + 2 + 2

ii Original marginal subtotals. Calculate adjustment for sex:

Subtotals	Means	Adjustments	Subtotals	Subtotals
			572	734
1982	86·17	−18	748	819
886	68·15	0	733	713
			815	602
2868			2868	2868

iii Adjust all subtotals and calculate adjustments for diet:

Subtotals	Sub-totals	Means	Adjust-ments	Subtotals
	464	51·56	+20	608
1568	640	71·11	0	711
886	643	71·44	0	605
	707	78·56	−7	530
2454	2454			2454

iv Adjust all subtotals and calculate adjustments for litter:

Subtotals	Subtotals	Sub-totals	Means	Adjust-ments
	644	654	65·40	+14
1646	640	730	66·36	+13
925	643	631	78·88	0
	644	556	79·43	0
2571	2571	2571		

Table 4 (*continued*): *Second cycle and final adjustments*

Sexes			Diets			Litters		
Subtotals	Means	Adjustments	Sub-totals	Means	Adjust-ments	Sub-totals	Means	Adjust-ments
v Adjust subtotals for litter and calculate adjustments for sex:								
			712			794		
1822	79·22	+0·16	721			873		
1032	79·38	0	710			631		
			711			556		
2854			2854			2854		
vi Adjust subtotals for sex and calculate adjustments for diet:								
			712·96	79·22	+1·00	795·12		
1825·68			721·96	80·22	0	873·96		
1032			710·80	78·98	+1·24	631·96		
			711·96	79·11	+1·11	556·64		
2857·68			2857·68			2857·68		
vii Adjust subtotals for diet and calculate adjustments for litter:								
			721·96			802·82	80·28	−0·28
1844·54			721·96			883·01	80·27	−0·27
1043·29			721·96			638·66	79·83	+0·17
			721·95			563·34	80·48	−0·48
2887·83			2887·83			2887·83		
viii Adjust subtotals for litter and calculate final adjustments for all factors:								
			719·96	79·996	0	800·02	80·002	0
1840·06	80·003	−0·003	720·17	80·019	−0·023	880·04	80·004	−0·002
1040·00	80·000	0	719·97	79·997	−0·001	640·02	80·002	0
			719·96	79·996	0	559·98	79·997	+0·005
2880·06			2880·06			2880·06		

ix Adjust grand total and calculate general effect:

Grand total (from *viii*)	2880·06
Adjustment for sex	−0·069
Adjustment for diet	−0·216
Adjustment for litter	+0·013
Final grand total	2879·788
General effect (i.e. mean)	79·99411

Table 5. *Collection of adjustments and computation of sum of squares*

Factor	First cycle	Second cycle	Final (*viii*)	Total adjustment	Effect	Subtotals (*ii*)
Sex M.	(*ii*) −18	(*v*) +0·16	−0·003	−17·843	+17·843	1982
F.	0	0	0	0	0	886
Diet A	(*iii*) +20	(*vi*) +1·00	0	+21·000	−21·000	572
B	0	0	−0·023	− 0·023	+ 0·023	748
C	0	+1·24	−0·001	+ 1·239	− 1·239	733
D	− 7	+1·11	0	− 5·890	+ 5·890	815
Litter I	(*iv*) +14	(*vii*) −0·28	0	+13·720	−13·720	734
II	+13	−0·27	−0·002	+12·728	−12·728	819
III	0	+0·17	0	+ 0·170	− 0·170	713
IV	0	−0·48	+0·005	− 0·475	+ 0·475	602
Note. The small roman numerals indicate from where the figures have been drawn				(*ix*)	+79·99411 General effect	2868 Grand total

Sum of products of effects and corresponding subtotals = 236,355

X^2/n = 228,484

Sum of squares for effects of factors = 7,871

Table 6. *Third analysis of variance*

Source of variation	Degrees of freedom	Sum of squares	Mean square	Variance ratio	Significance
Effects of the 3 factors	7	7871	1124	22·5	***
Interactions	19	1401	73·7	1·48	*n.s.*
Between 'combinations'	26	9272			
Residual	9	448	49·8		
Total	35	9720			

record the effects of these adjustments on all the subtotals. Thus the new male subtotal will be

$$1982 - 23(18) = 1568.$$

The female subtotal will remain unaltered. The subtotal for diet A is based on 6 males (and 3 females) as is shown in (*i*) of Table 4, and will therefore become

$$572 - 6(18) = 464.$$

Similarly, the subtotal for litter I is based on 7 males (and 3 females) and will therefore become

$$734 - 7(18) = 608.$$

The adjusted subtotals are set out in (*iii*) of Table 4. It is noted that agreement between the totals of the three sets provides a complete check on the adjustment of the subtotals.

Next the four diet means are computed:

$$51{\cdot}56 \quad 71{\cdot}11 \quad 71{\cdot}44 \quad 78{\cdot}56.$$

Since two significant figures are sufficient it will be convenient to make no adjustment for diets B and C, add 20 for diet A and subtract 7 for diet D:

$$51{\cdot}56+20 = 71{\cdot}56, \quad 71{\cdot}11, \quad 71{\cdot}44, \quad 78{\cdot}56-7 = 71{\cdot}56.$$

These adjustments are carried through into all subtotals as shown in (*iv*) of Table 4. Thus the new male subtotal will be

$$1568+6(20)+6(0)+5(0)+6(-7) = 1646.$$

The means for the four litters and hence the adjustments in respect of litter are now computed, as in (*iv*) of Table 4, and the adjustments made in (*v*). This completes the first cycle, adjustments having been made in turn for

sex—diet—litter.

The cycle of adjustments (carried to two decimals) may now be run through a second time as set out in (*v*), (*vi*) and (*vii*) of Table 4.

The process described will converge fairly rapidly to a solution, i.e. we shall find that after a few cycles the adjustments have become negligibly small. However, we can judge from the drop in the magnitude of the adjustments from the first cycle to the second that the third cycle adjustments will all be small. The third cycle can therefore be abridged, as in (*viii*) of Table 4, by calculating all means and all final adjustments, to three decimals, at once.

The subtotals, modified by these final adjustments, are not required, but the grand total should be adjusted as in (*ix*) of Table 4. The mean, found on dividing this final grand total by $n = 36$, will be the estimate of E, the so-called 'general effect'. It should be calculated to one or two decimals more than are provided in the final adjustments.*

In carrying out the adjustments as just described, it is, of course, immaterial on which factor we start or in which order we pass round the cycle, but other things being equal, it will usually be best to start with the factor producing the biggest effects and work down to the one with least effect. It is also immaterial what mean we aim at when applying adjustments, but the arithmetic is generally simpler if we aim at either the largest or smallest of the set so that one of the adjustments (in each set) is zero while the others are all of the same sign. However, in the first adjustment for diet ((*iii*) of Table 4) advantage was taken of the fact that the two *intermediate* values were by chance approximately the same (71·11 and 71·44), and *two* adjustments were thus made zero with a resultant slight simplification of the arithmetic.

The adjustments should now be collected together. For clarity this process has been set out in Table 5, but usually they could be picked up and added straight from Table 4. The resultant total adjustments, *with their signs changed*, will then be estimates of the 'effects', i.e. of the constants E_S, E_D and E_L. They may be reckoned to be correct or nearly correct to three decimals.

* The reason is that the sum of squares to be subsequently computed is near its maximum value (the residual sum of squares being near a minimum). The sum of squares is therefore relatively insensitive to errors in the estimated constants. Hence even if the estimates are good, say, to only three decimals, they may be treated, *for the purpose of calculating a sum of squares*, as though they were good to four or five decimals.

In the last column of Table 5 are written the *original* marginal subtotals against the corresponding effects and, at the bottom, the *original* grand total against the *final* adjusted mean (i.e. general effect). The sum of squares for the effects of the three factors is found by adding the products of the last two columns, not omitting the general effect × grand total, and subtracting the usual term, $X^2/n = (2868)^2/36$:

$$(17{\cdot}843)\,(1982) + (0)\,(886) + \ldots + (79{\cdot}99411)\,(2868) - 228{,}484 = 7871.$$

The number of degrees of freedom to be attributed to this sum of squares is 7, for, although 10 effects have been estimated, any arbitrary constant could be added to the effects in any one of the 3 sets (sex, diet and litter) and taken off the general effect without affecting the result, whence the number of degrees of freedom is

$$10 - 3 = 7.$$

Notice that sums of squares are not obtained individually for the three factors but only one sum of squares for the three acting together. The sum of squares, 7871, is not greatly different from that obtained in the approximate analysis of variance, 7923 (see Table 3), which suggests that the first four lines of the approximate analysis are all fairly reliable.

Subtracting the sum of squares for the effects (with 7 degrees of freedom) from the sum of squares between all combinations of the 3 factors (26 degrees of freedom) we are left with a sum of squares with 19 degrees of freedom representing various undifferentiated interactions between the factors (see Table 6). The corresponding mean square is larger, but not significantly larger, than the residual mean square, and hence there is at first sight no good evidence for interaction. However, the approximate analysis of variance (Table 3) did suggest that there is a significant interaction between sex and diet, and it is conceivable that this interaction, with only 3 degrees of freedom, has been hidden by being diluted with other interactions. It is therefore proper to make an exact analysis to test the significance of the sex × diet interaction.

If no restrictions are placed on the form of the interaction between sex and diet, then there will be an 'effect' appropriate to *every* combination of these two factors. This means that together they constitute virtually a single factor whose 'levels' are specified by the combination, e.g. male on diet A, female on diet C, etc. The data can therefore be treated as though they had arisen from a bi-factorial experiment.

The procedure for calculating the adjustments, as set out in Table 7, does not differ greatly from that employed with three factors. Sex × diet is now a single factor, and it is only to save paper that the 8 combinations have been laid out in a 4 × 2 table. It is found convenient to display the 4 litters horizontally instead of vertically, so that the block of numbers under litter I, etc., in (*i*) of Table 7 have the right pattern and orientation for calculating the effect of the sex × diet adjustments on the litter subtotal. Thus having found the adjustments for the eight sex × diet combinations in (*ii*) of Table 7, we calculate the adjusted subtotal for litter I as follows:

$$734 + \begin{matrix} 2(0) & 1(0) \\ -2(27) & -1(5) \\ -1(29) & -1(4) \\ -2(36) & 0(9) \end{matrix} = 570$$

The convergence is not so rapid as in Table 4, and the third decimal is therefore of doubtful value.

The sum of squares for effects (calculated in Table 8) has 10 degrees of freedom since there are now *two* sets of effects, one with 8 and one with 4, and

$$8+4-2=10.$$

Alternatively we may argue that there are 7 independent comparisons between the 8 sex × diet combinations and 3 independent comparisons between the 4 litters. Hence number of degrees of freedom is

$$7+3=10 \quad \text{as before.}$$

The sum of squares (7 degrees of freedom) for the three independent factors previously obtained may now be subtracted from this sum of squares with 10 degrees of freedom to leave a sum of squares with 3 degrees of freedom representing the interaction between sex and diet, as set out in Table 9. The balance of 16 degrees of freedom to make up the 26 degrees of freedom between all combinations represents undifferentiated interactions between sex and litter, diet and litter and between all three factors.

From the analysis of variance (Table 9) it appears that not only are the other interactions non-significant, but their mean square is scarcely greater than the residual mean square. The variance ratio for the sex × diet interaction is

$$F = 196{\cdot}0/49{\cdot}8 = 3{\cdot}936,$$

and is therefore just significant at the 5 % level. There is accordingly some moderately good evidence of an interaction between sex and diet.

We may now examine the nature of this interaction by listing the differences, male *minus* female, on the four diets:

Diet	Effects (from Table 8)		Difference M. − F.
	Male	Female	
A	+ 3·090	0	+ 3·090
B	+29·757	+ 8·933	+20·824
C	+29·100	+ 7·590	+21·510
D	+37·800	+10·353	+27·447

The table gives the impression that diet A may be the exception. For A the responses of the two sexes are not greatly different, while for B, C and D there is a big difference between the two sexes, but this difference may possibly not vary significantly among the three diets. On the other hand, the data also suggest the alternative hypothesis that the degree of difference between the sexes rises with the response, i.e. that the better the diet, the more the males will outstrip the females. It is very necessary to remember that the same data may be consistent with several hypotheses which, biologically considered, are fundamentally different. It is legitimate, and indeed desirable, to employ *a priori* biological knowledge at this stage even though the knowledge may amount to little more than a hint that one hypothesis may be more 'reasonable' than another. Often, however, no *a priori*

Table 7. *Calculation of adjustments for two factors, sex × diet and litter (first two cycles)*

i Composition of the subtotals:

	Sex × diet combinations: Male	Sex × diet combinations: Female	Litters: I	Litters: II	Litters: III	Litters: IV
			10	11	8	7
			=	=	=	=
A	6 = 2 + 2 + 1 + 1	3 = 1 + 0 + 1 + 1	2 + 1	2 + 0	1 + 1	1 + 1
B	6 = 2 + 2 + 1 + 1	3 = 1 + 1 + 1 + 0	2 + 1	2 + 1	1 + 1	1 + 0
C	5 = 1 + 1 + 2 + 1	4 = 1 + 2 + 0 + 1	1 + 1	1 + 2	2 + 0	1 + 1
D	6 = 2 + 1 + 2 + 1	3 = 0 + 2 + 0 + 1	2 + 0	1 + 2	2 + 0	1 + 1

ii Original marginal subtotals. Calculate adjustments for sex × diet:

Subtotals		Means		Adjustments		Subtotals			
381	191	63·50	63·67	0	0	734	819	713	602
541	207	90·17	69·00	− 27	− 5				
462	271	92·40	67·75	− 29	− 4				
598	217	99·67	72·33	− 36	− 9				
2868				Totals		2868			

iii Adjust all subtotals and calculate adjustments for litter:

381	191	Subtotals	570	669	551	497
379	192	Means	57·00	60·82	68·88	71·00
317	255	Adjustments	+ 14	+ 10	+ 2	0
382	190					
2287		Totals	2287			

iv Adjust all subtotals and calculate adjustments for sex × diet:

431	207	71·83	69·00	− 2·83	0	710	779	567	497
429	218	71·50	72·67	− 2·50	− 3·67				
345	289	69·00	72·25	0	− 3·25				
424	210	70·67	70·00	− 1·67	− 1·00				
2553				Totals		2553			

v Adjust all subtotals and calculate adjustments for litter:

414·02	207·00	Subtotals	689·08	754·50	554·66	485·75
414·00	206·99	Means	68·91	68·59	69·33	69·39
345·00	276·00	Adjustments	+ 0·48	+ 0·80	+ 0·06	0
413·98	207·00					
2483·99		Totals	2483·99			

Table 7 (*continued*): *Final adjustments*

vi Adjust all subtotals and calculate final adjustments for both factors:

Sex × diet combinations						Litters			
416·64	207·54	69·440	69·180	−0·260	0	693·88	763·30	555·14	485·75
416·62	208·33	69·437	69·443	−0·257	−0·263	69·388	69·391	69·392	69·393
346·40	278·08	69·280	69·520	−0·100	−0·340	+0·005	+0·002	+0·001	0
415·86	208·60	69·310	69·533	−0·130	−0·353				
2498·07				Totals		2498·07			

vii Adjust grand total and calculate general effect:

Grand total (from *vi*)	=	2498·07
Adjustments for sex × diet	=	−7·590
Adjustments for litter	=	+0·080
Final grand total	=	2490·560
General effect (i.e. mean)	=	69·18222

Table 8. *Collection of adjustments and calculation of sum of squares*

Factor		First cycle	Second cycle	Final (*vi*)	Total adjustment	Effect	Original subtotals
Combinations of sex and diet	M., A	(*ii*) 0	(*iv*) −2·83	−0·260	−3·090	+3·090	381
	M., B	−27	−2·50	−0·257	−29·757	+29·757	541
	M., C	−29	0	−0·100	−29·100	+29·100	462
	M., D	−36	−1·67	−0·130	−37·800	+37·800	598
	F., A	0	0	0	0	0	191
	F., B	−5	−3·67	−0·263	−8·933	+8·933	207
	F., C	−4	−3·25	−0·340	−7·590	+7·590	271
	F., D	−9	−1·00	−0·353	−10·353	+10·353	217
Litters	I	(*iii*) +14	(*v*) +0·48	+0·005	+14·485	−14·485	734
	II	+10	+0·80	+0·002	+10·802	−10·802	819
	III	+2	+0·06	+0·001	+2·061	−2·061	713
	IV	0	0	0	0	0	602
General effect and grand total (*vii*)						+69·18222	2868

Sum of products of effects and corresponding subtotals = 236,943

X^2/n = 228,484

Sum of squares for effects of factors = 8,459

Table 9. *Fourth analysis of variance*

Source of variation	Degrees of freedom	Sum of squares	Mean square	Variance ratio	Significance
Three independent factors	7	7871	1124	22·6	***
Sex × diet interaction	3	588	196·0	3·936	*
Three factors and all sex × diet interactions	10	8459			
Other interactions	16	813	50·8		*n.s.*
Between all combinations	26	9272			
Residual	9	448	49·8		
Total	35	9720			

knowledge is relevant, and in such cases the greatest care should be taken to avoid adopting an hypothesis merely because it is not belied by the data.

It is, of course, legitimate to test whether the data do or do not contradict *any hypothesis*, always remembering that failure to contradict the hypothesis is not a proof that the hypothesis is true. Hence, although it is not necessarily true that the most appropriate hypothesis to test is that diet A is exceptional, there can be no objection to making such a test as an illustration of the statistical technique.

The hypothesis to be examined is equivalent to supposing that one particular degree of freedom out of the three degrees of freedom for sex × diet interaction has a real effect while the other two have no effect. Hence we should separate the mean squares for the one and the two degrees of freedom respectively and endeavour to show that the former is and the latter is not significant.

Any algebraic formulation of the process would be cumbersome, but the arithmetic proceeds quite straightforwardly once the pattern of adjustments is appreciated (see Table 10). We are now permitting a different sex difference in diet A on the one hand and in diets B, C and D on the other. Hence the adjustment for sex is made separately in the two groups of diets (A) and (B, C or D). To avoid bandying adjustments backwards and forwards between sexes and diets, it is advisable to adjust only the male results. The means of the two sexes on diet A only are

$$\begin{array}{ll} \text{Male} & 381/6 = 63{\cdot}50 \\ \text{Female} & 191/3 = 63{\cdot}67 \end{array}$$

These are sufficiently close to make no adjustment necessary on the first cycle. The sex means calculated from the pooled data for diets B, C and D are

$$\begin{array}{lr} \text{Male} & 1601/17 = 94{\cdot}18 \\ \text{Female} & 695/10 = 69{\cdot}50 \\ \text{Difference} & = 24{\cdot}68 \end{array}$$

Hence adjustment is made by subtracting 25 from every male on diets B, C or D, leaving unaltered both males and females on diet A and females on diets B, C and D.

The adjustments to the subtotals are made in (*iii*) of Table 10. Thus for diets

$$\text{A} \quad 572 - 6(0) \;\; = 572 \text{ etc.}$$

and for litters

$$\text{I} \quad 734 - 5(25) = 609 \text{ etc.}$$

The final adjustments shown in (*viii*) of Table 10 are appreciable, so the third decimal is of doubtful value.

The number of degrees of freedom in the sum of squares for effects is now 8, being made up of 1 for sex in diet A, 1 for sex in diets B, C and D, 3 for differences between diets and 3 for differences between litters:

$$1 + 1 + 3 + 3 = 8.$$

The sum of squares with 7 degrees of freedom obtained with three independent factors (see Table 6) may be subtracted from the sum of squares with 8 degrees of freedom, to leave a square of 1 degree of freedom which measures the effect of the specific interaction which we have selected for examination. The balance to bring the sum of squares with 8 degrees of freedom up to the sum of squares with 10 degrees of freedom found when all interactions between sex and diet were permitted (see Table 9) is a sum of squares with 2 degrees of freedom and measures the effect of the interaction between sex and diet *other than* the specific degree of freedom which we selected. The full analysis of variance is set out in Table 12.

It is clear that the hypothesis examined is acceptable on these data. The selected degree of freedom in interaction is fairly significant, whereas the mean square for the remaining 2 degrees of freedom is neither significant nor even much greater than the residual sum of squares. In judging the significance of the selected degree of freedom one should, of course, allow for the fact that it was chosen as being the biggest of three. However, the three together were shown to be on the border-line of significance, so there is little danger that we have misled ourselves by selecting a square which is large only by chance.

6. Presentation of conclusions

Since litter is independent of the other two factors, any comparison between sexes, between diets or between any combinations of sex and diet may be regarded as generally true of all the litters in this experiment.

There is some evidence that sex and diet are not independent factors, i.e. that the two sexes react differentially to the diets. This evidence only just passes the conventional 5 % level of significance, so the possibility must not be ruled out that an interaction of the observed magnitude has arisen fortuitously. It is impossible, on statistical evidence alone, to specify at all closely the nature of this interaction, for it is consistent with either of the following hypotheses or indeed with many others:

(*a*) That the difference between male and female is the same on diets B, C and D but different on A.

(*b*) That the sex difference increases steadily with the response of either sex, i.e. that the 'better' the diet, the more the male will outstrip the female.

If the decision to test either of these hypotheses had been made before examining the data, the significance of the particular degree of freedom examined would have been considered reasonably strong. Of course when, as in the present case, the hypothesis is suggested by the data themselves, one must be prepared to discount to some extent the impression given by the moderately high level of significance.

Table 10. *Calculation of adjustments for three factors with one degree of freedom in interaction*

i Composition of subtotals:

Sexes × diets	Diets	Litters
Diet A (litters)	(litters) (sexes)	(diets)
M. 6 = 6 = 2 + 2 + 1 + 1	A 9 = 3 + 2 + 2 + 2 = 6 + 3	I 10 = 3 + 3 + 2 + 2
F. 3 = 3 = 1 + 0 + 1 + 1		2 + 5 = 7 males
	B 9 = 3 + 3 + 2 + 1 = 6 + 3	II 11 = 2 + 3 + 3 + 3
		2 + 4 = 6 males
Diets B + C + D	C 9 = 2 + 3 + 2 + 2 = 5 + 4	III 8 = 2 + 2 + 2 + 2
		1 + 5 = 6 males
M. 17 = 6 + 5 + 6 = 5 + 4 + 5 + 3		
F. 10 = 3 + 4 + 3 = 2 + 5 + 1 + 2	D 9 = 2 + 3 + 2 + 2 = 6 + 3	IV 7 = 2 + 1 + 2 + 2
		1 + 3 = 4 males

ii Original marginal subtotals. Calculate adjustments for sex × diet:

Sub-totals	Means	Adjust-ments	Sub-totals	Means	Adjust-ments	Sub-totals	Means	Adjust-ments
381	63·50	0	572			734		
191	63·67	0	748			819		
1601	94·18	−25	733			713		
695	69·50	0	815			602		
2868			2868			2868		

iii Adjust all subtotals and calculate adjustments for diet:

Sub-totals	Means	Adjust-ments	Sub-totals	Means	Adjust-ments	Sub-totals	Means	Adjust-ments
381			572	63·56	+10	609		
191			598	66·44	+7	719		
1176			608	67·56	+6	588		
695			665	73·89	0	527		
2443			2443			2443		

iv Adjust all subtotals and calculate adjustments for litter:

Sub-totals	Means	Adjust-ments	Sub-totals	Means	Adjust-ments	Sub-totals	Means	Adjust-ments
441			662			672	67·20	+14
221			661			778	70·73	+10
1248			662			634	79·25	+2
740			665			566	80·86	
2650			2650			2650		

Table 10 (*continued*)

Sub-totals	Means	Adjust-ments	Sub-totals	Means	Adjust-ments	Sub-totals	Means	Adjust-ments
v Adjust all subtotals and calculate adjustments for sex × diet:								
491	81·83	−2·83	728			812		
237	79·00	0	737			888		
1368	80·47	+1·53	724			650		
820	82·00	0	727			566		
2916			2916			2916		
vi Adjust subtotals for sex × diet and calculate adjustments for diet:								
474·02			711·02	79·00	+3·91	813·99		
237·00			746·18	82·91	0	888·46		
1394·01			731·65	81·29	+1·62	654·82		
820·00			736·18	81·80	+1·11	567·76		
2925·03			2925·03			2925·03		
vii Adjust all subtotals and calculate adjustments for litter:								
497·48			746·21			831·18	83·12	+0·39
248·73			746·18			904·47	82·22	+1·29
1408·77			746·23			668·10	83·51	0
829·81			746·17			581·04	83·01	+0·50
2984·79			2984·79			2984·79		
viii Adjust all subtotals and calculate final adjustments for all factors:								
501·34	83·557	−0·350	750·96	83·440	+0·102	835·08	83·508	+0·007
249·62	83·207	0	751·72	83·524	+0·018	918·66	83·515	0
1417·38	83·375	+0·429	751·88	83·542	0	668·10	83·512	+0·003
838·04	83·804	0	751·82	83·536	+0·006	584·54	83·506	+0·009
3006·38			3006·38			3006·38		

ix Adjust grand total and calculate general effect:

Grand total (from *viii*)	3006·38
Adjustment for sex × diet	+5·193
Adjustment for diet	+1·134
Adjustment for litter	+0·157
Final grand total	3012·864
General effect (i.e. mean)	83·69067

Table 11. *Collection of adjustments and calculation of sum of squares*

Factor	First cycle	Second cycle	Final (*viii*)	Total adjustment	Effect	Subtotals (*ii*)
Sex in diet A						
M.	(*ii*) 0	(*v*) −2·83	−0·350	−3·180	+3·180	381
F.	0	0	0	0	0	191
Sex in diets B, C and D						
M.	−25	+1·53	+0·429	−23·041	+23·041	1601
F.	0	0	0	0	0	695
Diet A	(*iii*) +10	(*vi*) +3·91	+0·102	+14·012	−14·012	572
B	+7	0	+0·018	+7·018	−7·018	748
C	+6	+1·62	0	+7·620	−7·620	733
D	0	+1·11	+0·006	+1·116	−1·116	815
Litter I	(*iv*) +14	(*vii*) +0·39	+0·007	+14·397	−14·397	734
II	+10	+1·29	0	+11·290	−11·290	819
III	+2	0	+0·003	+2·003	−2·003	713
IV	0	+0·50	+0·009	+0·509	−0·509	602
				General effect (*ix*)	+83·69067	
				Grand total (original)		2868

Sum of products of effects and corresponding subtotals = 236,817

X^2/n = 228,484

Sum of squares for effects of factors = 8,333

Table 12. *Fifth and final analysis of variance*

Reference	Source of variation	Degrees of freedom	Sum of squares	Mean square	Variance ratio
T. 6	Three independent factors	7	7871	1124	22·6***
Dif.	Selected sex × diet interaction	1	462	462	9·3*
T. 11	3 factors and a selected interaction	8	8333		
Dif.	Remaining sex × diet interactions	2	126	63	*n.s.*
T. 9	3 factors and sex × diet interactions	10	8459		
T. 9	Other interactions	16	813	50·8	*n.s.*
T. 9	Between all factorial combinations	26	9272		
T. 9	Residual	9	448	49·8	
	Total	35	9720		

If we accept the evidence for interaction between sex and diet, the results of the experiment may be summarized by the estimated effects shown in Table 8. The magnitude of the differences between litters are, however, of little interest. Each combination of sex and diet may therefore be averaged over the four litters. We have

$$69{\cdot}182 - \tfrac{1}{4}(14{\cdot}485 + 10{\cdot}802 + 2{\cdot}061 + 0) = 62{\cdot}345.$$

To this we may add the effect for each diet × sex combination; thus for males on diet A we have

$$62{\cdot}345 + 3{\cdot}090 = 65{\cdot}435.$$

It is unnecessary to quote the decimal figures; the calculation was carried to such accuracy only to illustrate the arithmetical processes. The conclusions can therefore be summarized in the following table:

Expected gains in weight (*g.*)

Diet	Male	Female	M. − F.
A	65	62	3
B	92	71	21
C	91	70	21
D	100	73	27

This table, read in conjunction with the notes on the significance of the sex × diet interaction, summarizes all the information in the experiment.

7. Discussion of methods

We have shown that even when orthogonality restrictions have completely broken down in a factorial experiment, it is still possible to construct exact analyses of variance and to make valid tests of the significance of any effects or interactions or in fact of any form of departure from any hypothesis. The reader may, however, have been left with an impression that the whole thing is a *tour de force* rather than a recognizable method of analysis. We regret if such is the impression; the calculations are long and tedious, but we hope that they follow a pattern of which the logic can be appreciated.

The principle of construction of a valid analysis of variance, in so far as it can be summed up in a few sentences, is as follows: Take any two hypotheses X and Y, of which X is admitted and Y is more complicated. Calculate the effects which are specified by X and the sum of squares for these effects. Do the same for hypothesis Y. Then the increases in the number of degrees of freedom and in the sum of squares as we pass from X to Y will lead to a mean square for deciding whether we must adopt the more complicated hypothesis. Thus in constructing the analysis of variance in Table 9 the relevant hypotheses were:

		Degrees of freedom
X	Sex, diet and litter have significant effects but none of their interactions is significant	7
	Difference	3
Y	Same as X, except that possibility of interaction between sex and diet is admitted	10

The difficulty which is peculiar to a non-orthogonal experiment is that the effects of irrelevant factors, such as litter in the example quoted, have to be computed separately on

the two hypotheses, X and Y. In an orthogonal experiment the effects of litters (and hence the corresponding sum of squares) would be unaltered by any modification of the hypothesis concerning the other two factors. This, when translated into general terms, explains why the statistical analysis of a non-orthogonal (and unbalanced) experiment will necessarily be troublesome; in fact, as we have already emphasized, the moral to be pointed is that the research worker should either acquaint himself with the principles of experimental design or else employ a statistician before embarking on the experiment.

SUMMARY

Data are presented from a biological experiment in which no attempt was made to ensure orthogonality or even balanced non-orthogonality. The proper design for an experiment with the same material is briefly discussed. The methods for the statistical analysis of the data are then explained by means of the arithmetical computations laid out in such a form that the routine can be understood and hence applied to other examples.

COMPARISONS OF HEIGHTS AND WEIGHTS OF GERMAN CIVILIANS RECORDED IN 1946–7 AND ROYAL AIR FORCE AND OTHER BRITISH SERIES

By G. M. MORANT, D.Sc., *R.A.F. Institute of Aviation Medicine*

1. Summary and main conclusions

1·1. Records of heights and weights collected during the war show clearly that changes were taking place in the distributions of the measurements for the British population. Compared with pre-war years maximum height, indicating skeletal maturity, was normally reached at a younger age than previously. Weights for the war period tended to be above pre-war levels for ages under about 25 years and below them for ages over 25. New standards were needed which could be used to assess the significance of contemporary and future changes in heights and weights. This paper is a contribution to that topic, which is aided by comparison of the recent British records with those for large numbers of civilians in the British zone of Germany measured for the Public Health Branch of the Control Commission for Germany (British Element) from December 1946 to June 1947.

1·2. Comparisons made lead to the following general conclusions:

(*a*) Appreciable differences are found between the mean heights, recorded weights and weights standardized for height of the German children of different towns (Hamburg and Berlin) and regions (*Länder*). For adults the only relation between the mean weights standardized for height which is consistent for both sexes and nearly all age groups is that the Berlin series are inferior to all others.

(*b*) Pooled series representing all regions except Berlin are treated in later comparisons. The data were subdivided to represent earlier (December 1946–February 1947) and later (March–June 1947) periods. None of the secular differences between the mean heights and weights of the German children, and between the weights of the adults, are large and it is possible that they were due partly to normal seasonal fluctuations which may be exhibited by a population having constant nutritional conditions.

(*c*) Occupational series of adults show a sequence in mean weights from very heavy workers (heaviest) to normal consumers (lightest).

(*d*) The total series, excluding Berlin children and adolescents, give growth curves for height which are of the usual forms, but which have maxima at the remarkably young ages of 19½ years for males and 17 years for females. For both sexes the weight curves are peculiar in showing less increase for ages over 18 than is normally found.

(*e*) For ages under 19 and comparisons of weights reduced at each age to a constant height, the post-war German standards occupy low positions, but ones which are not extremely low compared with those for pre-war German series.

(*f*) The maximum of the male age curve for height given by the post-war survey is greater than any of the pre-war means for regional sections of the male adult population of the British zone of Germany. These pre-war standards were probably underestimates of the mean heights of German men.

(*g*) The following conclusions ((*g*)–(*j*)) are derived from comparisons of mean weights for different series reduced to the same height at each age, i.e. for weights compared after

allowance has been made for differences in height. For adolescent male British industrial workers, series measured during the war (1943) were decidedly heavier than series measured before the war (1929–32). The British series of both dates are well above the post-war German series. For British female industrial workers, series measured during the war show a higher standard than pre-war series (1926), the latter being rather lower than the post-war German series.

(*h*) For men aged 17–40 years the mean standardized weights for twelve British and the post-war German series give the sequence: Commonwealth and British serving aircrew (heaviest)—aircrew recruits—R.A.F. ground staff—industrial workers, 1943—industrial workers, 1929–32—Germans—conscripts, including those rejected after examination, 1917–18—unemployed industrial workers, 1929–32 (lightest).

(*i*) The situation is markedly different for ages over 40. The German series clearly falls to the lowest place, and unlike all other series it shows mean weights rapidly declining with increasing age: the British wartime industrial workers fall below the pre-war employed workers, though remaining above the low levels of the 1917–18 conscripts and pre-war unemployed workers. Rationing has affected most markedly the weights of the middle aged and aged.

(*j*) The few series of women lead to similar conclusions. British industrial workers were decidedly heavier during than before the war, and the Germans were clearly at a lower level for all except younger ages (up to about 30 years). The evidence suggests that for all ages loss of weight due to restricted rations has been greater for females than for males, and decidedly greater in the case of both sexes for middle-aged and old people than for young adults and children.

2. THE GERMAN 1946–7 SURVEY

2·1. Starting in December 1946, the Public Health Branch of the Control Commission for Germany (British Element) has carried out a new survey of the body weights of German people in all parts of the British zone, including the British sector of Berlin. The plan was to remeasure the same people, as far as possible, at monthly intervals, the weights being taken during the first two weeks of each month. They were recorded under the direction of Public Health Officers, who aimed, in accordance with guidance received, at obtaining random samples of the populations of their districts. Both sexes and all age groups were to be represented. The returns were reduced by No. 1 Nutrition Survey Team and Hollerith machines were used for the purpose. The writer is indebted to Brigadier W. S. Martin, M.C., Public Health Adviser, Health Branch, C.C.G., and to F. D. G. Bailey, Esq., formerly Nutrition Officer of the same Branch, for permission to use the German data given in a new form in the present paper.

2·2. The records available consist of a preliminary report on the survey (Control Commission for Germany, 1947) with tabulated data for the months December 1946–March 1947, and additional tables for the months April–June 1947. Only average values of the measurements—heights and weights of children and adolescents and weights standardized for height of adults—are given. The total material can be divided into two sets, viz.

(i) Children and adolescents: (*a*) means of heights and weights as recorded for each month December 1946–February 1947 (and for some regions also for March 1947), distinguishing sexes, ages (denoted by year of birth), and Berlin, Hamburg and regional samples, the data

also being given for the total British zone except Berlin; (*b*) the same data for each month March–June 1947 for the total British zone except Berlin only.

(ii) Adults: (*a*) mean weights reduced to constant heights (see § 3·6) for each month December 1946–February 1947 (and for some regions also for March 1947), distinguishing sexes, ages (by decennial groups of years of birth), occupational groups, and Berlin, Hamburg and regional samples, the data also being given for the total British zone except Berlin; (*b*) the same data for each month March–June 1947 for the total British zone except Berlin only.

The age ranges for the children and adolescent series, on the one hand, and for the 'adult', on the other, overlap. The youngest group for the latter is 7–17 years and data for this were not used as it covers too large a part of the age cycle when growth is most rapid. Other details regarding the survey are discussed in § 3 below.

3. Treatment of the German 1946–7 survey data

3·1. The available statistical data for the German survey consist of means only. Possible ways of treating them are thus very limited and rigorous comparisons cannot be made. The aim of the comparisons in following sections of this paper is to reveal the salient characteristics of the material only. The treatment consists chiefly in deriving clear-cut conclusions from graphical comparisons. The means are for series distinguished on account of sex, age, locality, time (month of measurement) and occupation (adults only). By appropriate grouping of the means, providing pooled means, the situation is examined for each of the last three of these factors considered singly, possible interaction between them being ignored. This omission is not likely to have given invalid conclusions, since notice is only taken of marked distinctions.

3·2. The aim was to remeasure the same people at monthly intervals, so that information would be obtained most economically regarding a secular change in the distributions of weights for the populations sampled. For various reasons this was not achieved fully and it is said: 'The number of subjects who were actually weighed in consecutive months is much lower than might be expected.' The number of individuals giving a mean for any particular month is known. In pooling data for different months the number of observations is known but not the number of individuals involved, because the number of repeated weighings is unknown. It may be noted that the n's for means given in the tables below have greater statistical value in secular comparisons and lesser value in regional and occupational comparisons.

3·3. The original grouping is by calendar years of birth, not years of age. When translated into years of age, with regard to the date or dates of measurement, the central values of groups are fractions of years. In comparing the data for children with other series it is more convenient to have means for even central values of the form x·0 or x·5 years. Estimates at these ages were obtained for the German and other series when necessary by linear interpolation, which can be supposed sufficiently precise when data at 1-year intervals are being treated. Consideration of the following points is necessary.

3·4. *Seasonal fluctuations in height and weight increments*. The German data are for the months December 1946–June 1947, measurements having been recorded during the first two weeks of each month. It is convenient to make comparisons between pooled means for

the two periods December 1946–February 1947 and March–June 1947. In doing this, differences observed might be partly due to normal seasonal changes in growth rates for height and weight in the case of immature individuals, and to normal seasonal fluctuations in the weights of adults. The same question has to be taken into account in comparing the German series as a whole with others. Data regarding seasonal changes of heights and weights in children have been given by Schmid-Monnard (1895), Mumford (1927), Friend (1935) and Friend & Bransby (1947), and for adolescents and adults by Kemsley (1945–7). It has been found for schoolchildren that growth at all times of the year is greater during holidays than during term time. The evidence of different series is not entirely in agreement, but it justifies some rather vague general conclusions. These are that seasonal fluctuations in the growth rate of height in boys (and presumably girls, too) are probably so small that they may be neglected in comparing different series; that they may be appreciable for weights of children—giving maximum differences between quarterly averages for the same population of the order 2 lb.—and that they are probably negligible for weights of adults.

3·5. *Allowances for clothes.* In comparing the heights and weights of different series of children and adults, allowance has to be made for differences in the clothing worn by the subjects when measured. The allowances are required to adjust average values, and there can be no assurance that they are precisely correct when applied to a particular series. Relevant information is:

(*a*) With reference to a survey of male industrial workers: 'Height was taken without shoes, but with very thin heel-less slippers. Weights were taken with ordinary clothes on, 8 lb. being deducted from the gross weights registered, this value being the average of repeated weighings of male clothing' (Cathcart, Hughes & Chalmers, 1935).

(*b*) With reference to a survey of female industrial workers: 'Weight was taken without shoes, light heel-less slippers being provided.... Assessment was made of the weight of clothing worn by the average worker, and as a result 4 lb. was deducted from the gross weights obtained' (Cathcart, Bedale, Blair, Macleod & Weatherhead, 1927).

(*c*) The following means weights in kg. (1 kg. = 2·2 lb.) are given for the clothes of Alsatian boys measured in the spring (Schlesinger, 1917):

Age	7	8	9	10	11	12	13	14
Coat	0·2	0·2	0·25	0·3	0·4	0·5	0·6	0·6
'Trousers'	0·25	0·25	0·3	0·35	0·4	0·4	0·5	0·6
Underclothes	0·25	0·25	0·3	0·3	0·3	0·3	0·3	0·3
Boots	0·5	0·6	0·7	0·7	0·8	0·8	0·9	0·9
Total	1·2	1·3	1·55	1·65	1·9	2·0	2·3	2·4

(*d*) Measurements of R.A.F. aircrew clothed and unclothed (taken in December) give the following data:

Average excess of height in socks over height nude = 0·27 in. (520).

Average excess of weight clothed, without tunic or boots and with trouser pockets empty, over weight nude = 4·9 lb. (520).

Average weight of tunics = 2·5 lb.

Average weight of boots = 3·5 lb.

Measurements of the German 1946–7 survey are recorded for subjects in indoor clothing and without shoes. For various reasons some of the subjects were actually measured in other conditions, and the following corrections were applied to adjust such readings so that all records would be for subjects wearing the prescribed clothing (indoor clothing and without shoes):

Apparel	Correction for height (cm.)		Correction for weight (kg.)	
	Males	Females	Males	Females
Shoes	−2	−3	−1·0	−0·5
Overcoat and shoes	−2	−3	−3·5	−2·0
Heavy working dress	−2	—	−5·0	—
Religious sisters dress	—	−2	—	−3·0
Underwear only	—	—	+0·7	+0·7
None	—	—	+1·0	+1·0

It is not stated that these estimates were obtained (as average values) by actually measuring subjects wearing different sets of apparel, or by weighing garments. Most of them are in fairly close agreement with earlier estimates, but the additions to be made to weights when either underwear only or no clothing was worn are certainly too small.

In comparing different series mean heights and weights were adjusted when necessary by referring to the data given above (*a*)–(*d*) and choosing arbitrary corrections for clothing suitable to the circumstances. For children and adolescents the reduction is to measurements in indoor clothing and without shoes, so the German 1946–7 survey means are accepted as given. For weights of adults the reduction is to nude measurements, since most of the series give records for this condition, and the German means are adjusted accordingly. There is no assurance that allowances made for clothes are very close approximations to the unknown values in all cases, but errors on this account are not likely to be greater than about ¼ in. (6 mm.) for height and 1 lb. (0·45 kg.) for weight.

3·6. *Reduction of weights to constant heights.* In comparing the mean weights of two series it is necessary to make proper allowance for differences between their mean heights at different ages. This is best done for each age by using the regression coefficient of weight on height (giving increase in mean weight for unit increase in mean height), and reducing the mean weights of one of the series to values expected if its mean heights were those of the other series. Two sets of weights, one set estimated, for people having the same average heights are thus compared. Regression coefficients provided by the following British series were examined, all samples being for adequate numbers of individuals:

(*a*) male industrial workers, 1929–32, age groups 15 to over 60 years (Cathcart *et al.* 1935);

(*b*) female industrial workers, 1926, age groups 15 to over 50 years (Cathcart *et al.* 1927);

(*c*) Glasgow children divided into four groups representing different social grades, age groups 6–13 years (Elderton, 1914);

(*d*) East Sussex children, age groups 5–14 years (Dunstan, 1925);

(*e*) R.A.F. aircrew, 1944, age groups 20–40 years (unpublished).

The regression coefficients for these series are in good agreement, and there are clear differences between the values on account of sex, age and social class. In general they in-

crease with increase in mean weight. The age curves provided by the coefficients for each sex, and the relations between the curves for the two sexes, are very similar to typical age curves for weight (e.g. Fig. 8), female values only exceeding male for some adolescent ages. Series representing higher social classes tend to have greater coefficients for corresponding ages than series representing lower social classes, but differences of this kind are relatively small.

In practice it is not essential that regression coefficients used for the purpose considered should be derived from either of a pair of series compared, or that values used should be very close approximations to such ideal data. This is so because differences between the mean heights of the series at corresponding ages—the factor by which the regression

Table 1. *Data used in reducing weights to constant heights*

Age group (years): central values		6·5	7·5	8·5	9·5	10·5	11·5	12·5
Accepted regression coefficients, x (1 cm. $= x$ kg.)	Males	0·27	0·29	0·31	0·33	0·36	0·40	0·44
	Females	0·25	0·27	0·29	0·31	0·35	0·40	0·46
Heights (cm.) to which weights were reduced	Males	118	122	127	131	136	140	144
	Females	117	121	126	130	135	141	147
Age group (years): central values		**13·5**	**14·5**	**15·5**	**16·5**	**17·5**	**18·5**	**All adult ages**
Accepted regression coefficients, x (1 cm. $= x$ kg.)	Males	0·49	0·55	0·60	0·63	0·65	0·65	0·65
	Females	0·51	0·53	0·54	0·55	0·55	0·55	0·55
Heights (cm.) to which weights were reduced	Males	149	155	162	168	171	173	168·5
	Females	152	157	161	161	161	161	157·5

coefficients have to be multiplied—will nearly always be small. The reduced weights given in this paper were obtained by using a constant set of regression coefficients, these having been made up by balancing the evidence of the series listed above.

In comparisons with the German 1946–7 survey, the position is different for the immature series, on the one hand, and the adult on the other (see § 2·2). In the case of the former, absolute mean heights and weights are available for each year of age. It is convenient (at each age) to reduce the mean weights of all other series to the mean heights of the German series, these mean heights being given, together with the regression coefficients used, in

Table 1. In the case of the adults no absolute constants are available, but only mean weights reduced to the heights of 1685 mm. for men and 1575 mm. for women. The reduction is said to have been made for all age groups by applying the allowances (1 cm. in height) = (0·65 kg. in weight) for men, and (1 cm.) = (0·55 kg.) for women. These are presumably regression equations derived from correlation tables for height and weight, and they show good agreement with equations given by British series of adults. Accordingly, for all adult ages the German equations, and the heights to which the weights were reduced, were adopted. It may be noted that these heights (1685 mm. for men and 1575 mm. for women) are probably rather lower than the averages for German adults (cf. Fig. 7), so the reduced weights provided are probably rather less than the actual averages for the people. In comparisons where weights were adjusted the data given in Table 1 were used.

In view of the loss in precision—due chiefly to allowances which have to be made for clothes, and partly to the method of reduction to constant heights—no significance should be attributed to small differences between mean weights for different series. A difference greater than 2 lb., say, is almost certainly real, but a difference of 1 lb. may not be.

4. The heights and weights of Germans recorded in 1946–7

4·1. *Regional comparisons.* Comparisons can only be made for the months December 1946–March 1947 (children), or December 1946–February 1947 (adults). Pooled means were calculated for these periods, restricted to the years of age for which the data are most adequate, and for separate regions. The mean heights and weights for boys are plotted in Figs. 1 and 2, and the diagrams for girls give a similar impression of regional distinctions. Assuming that the order of variation was not exceptional, it can be estimated, by using standard deviations for other series, that most of the differences between the German means at corresponding ages are statistically insignificant, but some of the differences are almost certainly markedly significant. Taking weighted mean differences for ages between 6 and 16 years suggests the following order for both boys and girls:

North Rhine—Westphalia—Berlin—Hamburg—Hannover—Schleswig-Holstein.
(tallest) (shortest)

The weights as recorded give the similar orders:

Boys: North Rhine—Westphalia—Hamburg—Berlin—Hannover—Schleswig-Holstein.
(heaviest) (lightest)
Girls: North Rhine—Hamburg—Westphalia—Schleswig-Holstein—Hannover—Berlin.

The orders given by the mean weights are changed substantially when allowances are made for differences between the mean heights. The reduced weights, obtained by the method described in § 3·6, give:

Boys: Schleswig-Holstein—North Rhine—Westphalia—Hamburg—Hannover—Berlin.
(heaviest) (lightest)
Girls: North Rhine—Schleswig-Holstein—Hamburg—Westphalia—Hannover—Berlin.

For both boys and girls the differences between the extremes here are small (less than 1 kg.) for the younger ages and decidedly larger (several over 2 kg.) for ages over 12 years.

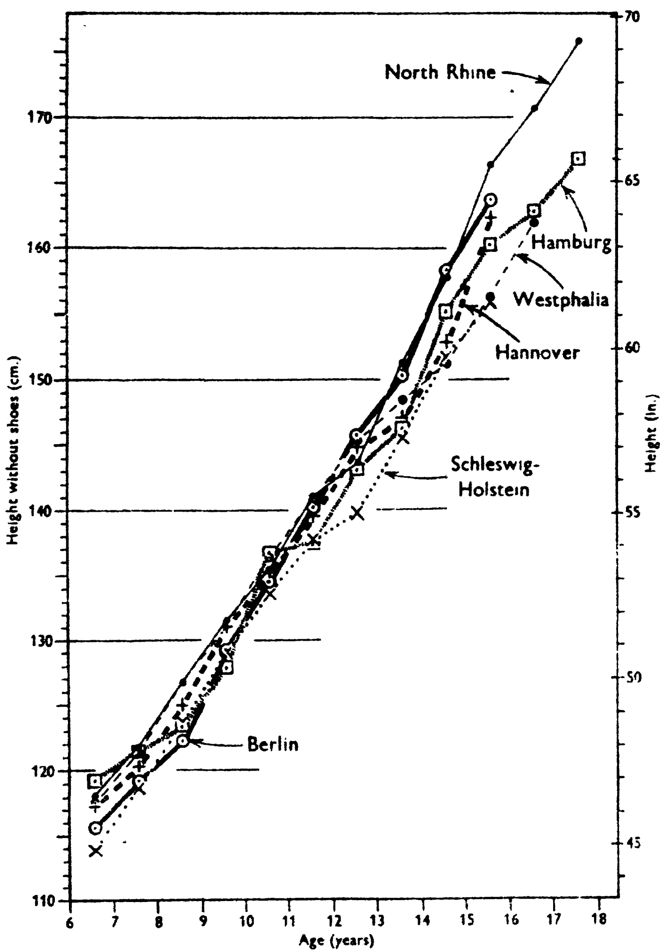


Fig. 1. Average heights for regional series of German boys, December 1946–March 1947.

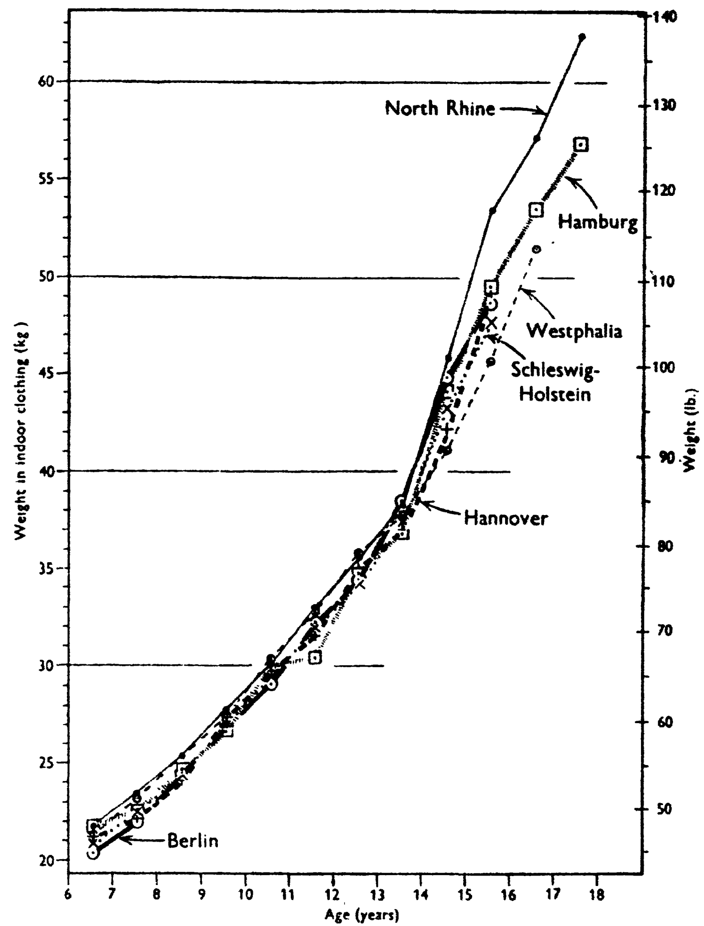


Fig. 2. Average weights (indoor clothing and without shoes) for regional series of German boys, December 1946–March 1947.

Reduced mean weights for the adults are shown in Fig. 3 and they suggest the following sequences:

Men: Hannover, Hamburg and Schleswig-Holstein—Westphalia—North Rhine—Berlin.
(heaviest and no appreciable differences between levels) (lightest)

Women: Westphalia—North Rhine, Hamburg and Schleswig-Holstein—Hannover—Berlin.
(no appreciable differences)

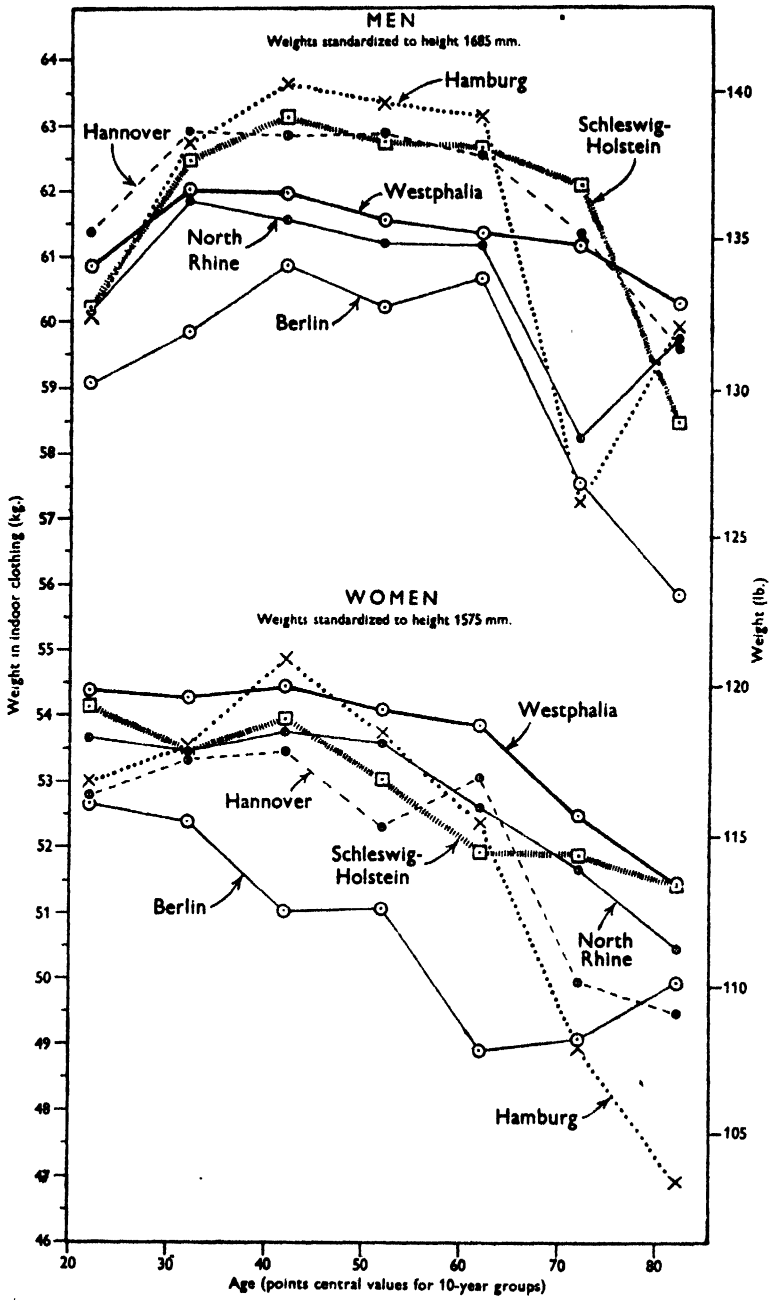


Fig. 3. Average weights (indoor clothing and without shoes) standardized for height of regional series of German men and women, December 1946–February 1947.

Excluding the Berlin series, the maximum differences between the reduced mean weights for the five other regions are of the order 2 kg. The only relation between the weights standardized for height which is consistent for both sexes and nearly all ages is that the people of Berlin were inferior to all the others. Regional differences between the other series are smaller in value and clearly of less significance.

4·2. *Secular comparisons: heights of children.* The total series considered here is that for the British zone of Germany excluding Berlin. The numbers of observations made in each month are roughly the same for the boys and girls and of the orders 10,000 for December 1946, 4000 for January 1947 and 11,000–13,000 for each of the months February–June 1947. A comparison of the mean heights for separate months failed to reveal any consistent secular trend. To obtain larger samples the seven months were divided into the two groups

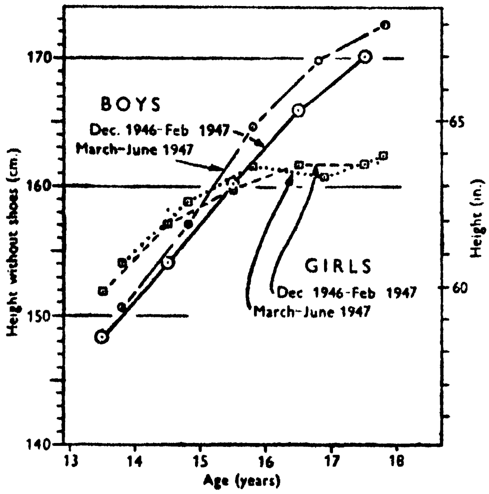


Fig. 4. Average heights of German boys and girls. (British zone except Berlin.)

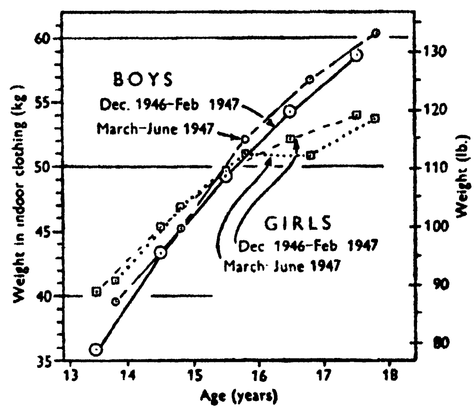


Fig. 5. Average weights (indoor clothing and without shoes) of German boys and girls. (British zone except Berlin.)

December 1946–February 1947 and March–June 1947, and the pooled means suggest the following conclusions:

Boys. Ignoring the two youngest ages for which the numbers are small, for ages under 13 years the two 'curves' are interlaced and the maximum divergence between them is about 15 mm. For ages over 13 (Fig. 4) the curves are separated and the divergence reaches a maximum of 25 mm. (1 in.) between ages 16 and 17 years, which is estimated to be markedly significant.

Girls. For ages under 13 the curves are interlaced and the maximum divergence is about 10 mm. For ages over 13 (Fig. 4) the means tend to be greater for the later period, but the maximum divergence is only 10 mm. and this may not be significant.

The general conclusion for both boys and girls is that there is no clear distinction between the growth rates for height of the population in the two periods for ages under 13; for ages over 13 there is indication, which is clear for boys and less clear for girls, that growth in height was faster in the later than in the earlier period. The distinction may be due partly to a normal seasonal fluctuation in the growth rate (see § 3·4).

4·3. *Secular comparisons: weights of children.* Comparisons of the age curves for the two periods leads to conclusions for weight similar to those for height. There are no clear distinctions for ages under 13. For ages over 13 (Fig. 5) the boys tend to be heavier for the

later period, the maximum divergence of the curves being about 1·4 kg. (3 lb.), but the girls tend to be practically the same or lighter for the later period. These comparisons relate to the actual weights recorded. Standardizing the means to the same height at each age, by the method described in § 3·6, has the effect of lessening the divergence of the weight curves and reversing the sign of the differences for most ages. The amounts in kg. by which the mean weights for the *earlier* period exceed those for the later after allowance has been made for differences in mean heights are:

Central value age group (years)	13·5	14·5	15·5	16·5	17·5
Boys	−2·4	+0·4	+0·8	+0·3	−0·1
Girls	+0·5	+0·5	+0·5	+0·9	+1·2

Interpreted in this way, the tendency was for children of the same age and height to be lighter in the later period, but the differences are small and of less significance because other evidence suggests that weight increments for children are normally rather greater during the winter than during the summer.

4·4. *Secular comparisons: weights of adults.* As for the children, comparisons are made between the two periods December 1946–February 1947 and March–June 1947. The data are mean weights standardized to constant heights. For males above age 40 the means are appreciably greater for the latter period: for females the same is true for ages over 70, but for ages about 30–70 the position is reversed. The distinction here is of more significance in view of the fact that weights of adults are believed to be normally rather greater in the winter than in the summer.

4·5. *Comparisons between occupational series.* The means are plotted in Fig. 6. The sequence: Very Heavy Workers (heaviest)—Heavy Workers—Miners—Moderately Heavy Workers and Normal Consumers (lightest) is clear for the men. For the women an order is less clear, but Normal Consumers tend to be appreciably lighter than Workers. The distinctions between the series must, of course, be influenced largely by processes of selection. For example, Very Heavy Workers (men) were as heavy for older as for younger age groups, which is not true for the total population (see Fig. 20). The distinction may be due to better feeding of the subgroup, or to the fact that middle-aged men below its average in physique and weight were liable to leave it because they were unable to continue heavy work, or both these factors may have been involved.

4·6. *Age curves for the total German series of children and adolescents.* (Table 3, Figs. 7, 8.) The series referred to are for the total British zone except Berlin. The age curves for height are almost coincident for ages under 11 years: for ages 11–15 the girls have the greater averages, this being the time when a clear puberty dip is shown by the curve for the boys. The curves provide evidence of the attainment of skeletal maturity at surprisingly young ages. The maximum average heights for males is seen to have been reached by 19·5 years, and for females by about 17 years. Extensive records for British men show that for the general population the maximum in question during the latter half of the nineteenth century was about 26 years, and that in the present century it tended to move to a younger age, being about 20 years to-day. The change is presumed to have been due to the better feeding of children. For some years before 1945 the German general population may have been fed better than the British, resulting in the attainment of maximum height, on the

Table 2. *Average weights taken at different periods of series of German men and women. Totals for British zone except Berlin.*

Years of birth	With indoor clothing and without footwear						Nude† (estimated)
	Months when measured						
	Dec. 1946–Feb. 1947		March–June 1947		Dec. 1946–June 1947		
	Mean (kg.)	c.v. age group	Mean (kg.)	c.v. age group	Mean (kg.)	c.v. age group	Mean (kg.)
	Men: weights standardized to height 1685 mm.						
1920–29	60·84 (13,382)*	22·0	61·00 (14,830)	22·3	60·92 (28,212)	22·2	57·4
1910–19	62·52 (11,357)	32·0	62·41 (12,472)	32·3	62·46 (23,829)	32·2	59·0
1900–09	62·44 (18,906)	42·0	62·89 (20,155)	42·3	62·67 (39,061)	42·2	59·2
1890–99	62·26 (11,908)	52·0	62·55 (12,992)	52·3	62·41 (24,900)	52·2	58·9
1880–89	62·07 (5,863)	62·0	62·33 (6,528)	62·3	62·21 (12,391)	62·2	58·7
1870–79	60·39 (2,140)	72·0	62·28 (2,544)	72·3	61·41 (4,684)	72·2	57·9
1860–69	59·68 (1,142)	82·0	61·12 (1,324)	82·3	60·45 (2,466)	82·2	56·9
1850–59	59·14 (71)	92·0	59·48 (52)	92·3	59·29 (123)	92·2	55·8
Totals	64,769	—	70,897	—	135,666	—	—
	Women: weights standardized to height 1575 mm.						
1920–29	53·67 (7,206)	22·0	53·41 (7,928)	22·3	53·53 (15,134)	22·2	51·5
1910–19	53·74 (9,763)	32·0	53·01 (10,899)	32·3	53·36 (20,662)	32·2	51·4
1900–09	53·50 (11,129)	42·0	53·30 (11,906)	42·3	53·40 (23,035)	42·2	51·4
1890–99	53·24 (8,239)	52·0	52·68 (8,564)	52·3	52·95 (16,803)	52·2	50·9
1880–89	52·60 (5,230)	62·0	51·91 (5,477)	62·3	52·25 (10,707)	62·2	50·2
1870–79	50·26 (3,738)	72·0	51·88 (4,317)	72·3	51·13 (8,055)	72·2	49·1
1860–69	49·80 (1,827)	82·0	51·27 (2,074)	82·3	50·58 (3,901)	82·2	48·6
1850–59	48·07 (208)	92·0	49·11 (202)	92·3	48·58 (410)	92·2	46·6
Totals	47,340	—	51,367	—	98,707	—	—

* A figure in brackets is the number of observations on which the mean is based and some repeated weights of the same people are included.
† Allowing for clothes 3·5 kg. (7·7 lb.) for men and 2 kg. (4·4 lb.) for women.

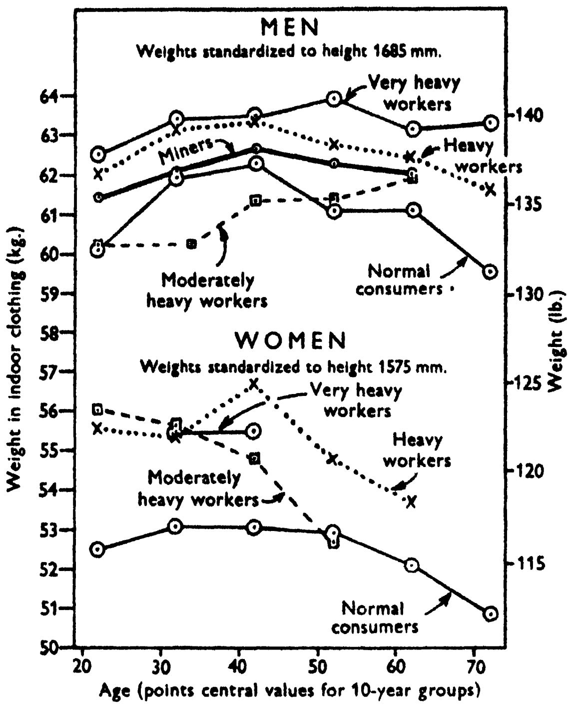


Fig. 6. Average weights (indoor clothing and without shoes) standardized for height of occupationa series of German men and women. (British zone except Berlin, December 1946–June 1947.)

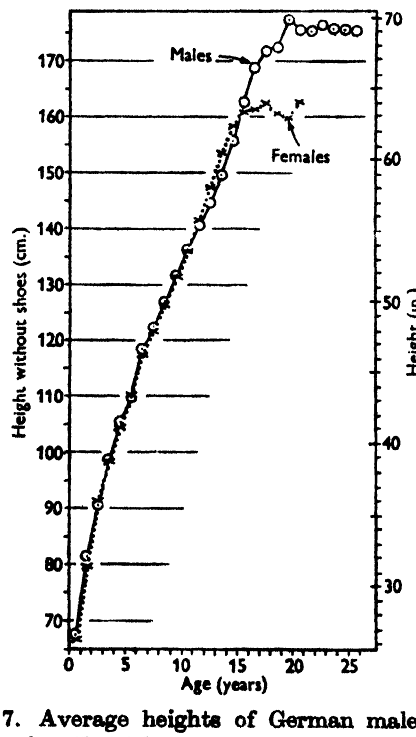


Fig. 7. Average heights of German males and females. (British zone except Berlin, December 1946–June 1947; data in Table 3.)

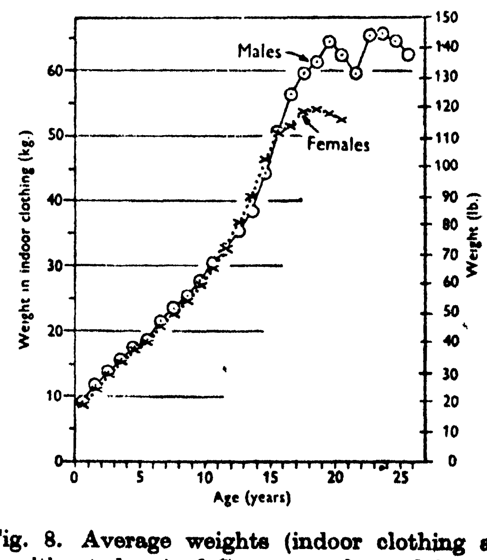


Fig. 8. Average weights (indoor clothing and without shoes) of German males and females. (British zone except Berlin, December 1946–June 1947; data in Table 3.)

Table 3. *Mean heights and weights of German boys and girls for the total British zone except Berlin*

A. *Measurements as recorded in indoor clothing and without shoes*

Age (years): central values	No. of measurements*		Mean height (mm.)		Mean weight (kg.)	
	Boys	Girls	Boys	Girls	Boys	Girls
0·7	153	135	676·7	668·3	9·03	8·49
1·7	116	71	814·9	798·6	11·77	11·11
2·7	648	448	907·7	914·7	13·94	13·26
3·7	1,055	1,076	985·5	984·1	15·84	15·34
4·7	1,451	1,407	1051·7	1040·7	17·54	17·12
5·7	1,673	1,444	1099·1	1101·2	18·87	18·29
6·7	3,714	3,112	1184·9	1174·6	21·67	20·79
7·7	8,868	8,143	1221·9	1219·5	23·23	22·39
8·7	8,390	8,569	1269·8	1264·9	25·35	24·55
9·7	9,523	9,420	1317·5	1315·0	27·83	27·01
10·7	10,879	11,414	1362·6	1360·2	30·28	29·74
11·7	10,246	11,048	1408·2	1413·7	32·81	32·83
12·7	8,752	9,057	1447·5	1472·1	35·71	36·66
13·7	5,167	4,896	1498·1	1534·1	38·28	40·92
14·7	3,277	3,076	1558·5	1581·0	44·48	46·23
15·7	1,750	1,436	1627·5	1609·2	50·95	50·48
16·7	834	547	1689·1	1610·6	56·20	51·32
17·7	452	416	1719·3	1621·5	59·87	53·81
18·7	119	216	1720·3	1606·2	61·25	54·04
19·7	74	57	1774·4	1598·2	64·56	53·24
20·7	33	32	1755·8	1622·7	62·40	52·18
21·7	19	—	1750·1	—	59·47	—
22·7	43	—	1763·0	—	65·27	—
23·7	37	—	1755·8	—	65·75	—
24·7	44	—	1753·9	—	64·58	—
25·7	43	—	1750·2	—	62·59	—
Totals	77,360	76,020	—	—	—	—

* These are the numbers of observations on which the means are based and some repeated measurements of the same individuals are included.

B. *Estimated nude weights reduced to height* 1685 *mm. for males and* 1575 *mm. for females*

Age (years): central values	Males†	Females‡
16·7	52·93 (834)	47·61 (547)
17·7	54·39 (452)	49·25 (416)
18·7	55·46 (119)	50·32 (216)
19·7	56·55 (74)	49·96 (57)

† Allowances for clothes 3 kg. (6·6 lb.) for age group 16·7, 3·25 for 17·7 and 3·5 for groups 18·7 and 19·7.

‡ Allowances for clothes 1·75 kg. (3·9 lb.) for age group 16·7, 2 for 17·7, 18·7 and 19·7.

average, at a still younger age. Restrictions thereafter would not be expected to modify the form of the age curve appreciably until after a lapse of some years.

The weight curves cross at about ages 11·5 and 15·5 years. Compared with pre-war data, both are peculiar in showing a marked decline in weight increments after age 18.

5. Comparisons of the post-war and earlier German series

5·1. Records of the heights and weights of German children living in the British zone of Germany are fairly abundant for years before 1939 and a few of the series were selected for comparison here. They are for:

(*a*) Boys and girls measured in Berlin about 1900, data for 'higher schools' (*Gymnasien*) and 'lower schools' (*Gemeindeschulen*) being given separately (Rietz, 1903).

(*b*) Boys and girls in a Berlin orphanage measured in 1919 (Davidsohn, 1919).

(*c*) Boys and girls (heights only) measured in 'lower schools' (*Volkschulen*) in Kiel and other towns in Holstein in 1902 (Ranke, 1905).

(*d*) Boys only measured in 'higher schools' in Hamburg about 1878 (Kotelmann, 1879, quoted by Ranke, 1905).

Means for these series are given in Tables 4 and 5 and those for the Berlin children are plotted, together with the means for the 1946–7 series, in Figs. 9–12.

5·2. As expected, the age curves for the pre-war Berlin series make clear distinctions between the social classes represented. It is surprising, however, to find that the curves for the general population in 1946–7 indicate a standard between those of the earlier 'higher school' children, on the one hand, and of the 'lower school' and orphanage children, on the other. This is so for both heights and weights of both boys and girls. The present-day Schleswig-Holstein series is also superior to earlier 'lower school' children of Holstein in the case of heights of boys and girls; and for Hamburg series the 1946–7 means exceed those of 'higher school' boys in 1878. The standards of the post-war German children are evidently not markedly depressed.

5·3. The average weights are best compared by reducing them at each age to a constant height (see § 3·6). The heights used for this purpose are those of the total 1946–7 series for the British zone except Berlin, and the weights of all the pre-war German series reduced to these heights are given in Table 7 and plotted in Figs. 17 and 18. In general the post-war series show standards of weights, after allowance is made for differences in height, which do not compare unfavourably with those of pre-war lower social grades of the German population (see also § 6·3).

5·4. The writer has been unable to find any records of large pre-war series of German adults suitable for comparison with the 1946–7 data. Maps compiled by anthropologists (Coon, 1939, and earlier authorities) show average heights for the male adult populations of different parts of the British zone of Germany ranging from 168 to 173 cm. The mean for the region in 1946–7 is about 176 cm. (Fig. 7). A probable explanation of the superiority of the later estimate is that it represents the maximum of the age curve (reached to-day about 19½ years), whereas the earlier estimates were derived mainly from data for conscripts (ages about 18–20) at times when the maximum of the age curve was normally not attained until some age near 25 years. It is known for the general British population that the age in question changed from about 26 years in the latter half of the nineteenth century to about 20 years to-day, while the maximum mean height attained remained unchanged in the past hundred years.

Table 4. *Average heights (without shoes) in mm. for pre-war series of German children*

Age (years): central values	Boys					Girls			
	Berlin			Holstein	Hamburg	Berlin			Holstein
	'Higher schools' *c.* 1900	'Lower schools' *c.* 1900	Orphanage 1919	'Lower schools' 1902	'Higher schools' *c.* 1878	'Higher schools' *c.* 1900	'Lower schools' *c.* 1900	Orphanage 1919	'Lower schools' 1902
6	—	—	988 (86)	—	—	—	—	1006 (74)	—
6·5	1183 (45)	1136 (128)	—	1059 (47)	—	1190 (14)	1119 (110)	—	1073 (43)
7	—	—	1068 (82)	—	—	—	—	1107 (45)	—
7·5	1220 (101)	1172 (189)	—	1117 (185)	—	1227 (42)	1173 (159)	—	1136 (81)
8	—	—	1142 (58)	—	—	—	—	1135 (55)	—
8·5	1273 (156)	1214 (198)	—	1169 (189)	—	1272 (37)	1217 (164)	—	1165 (108)
9	—	—	1175 (74)	—	—	—	—	1185 (47)	—
9·5	1312 (168)	1265 (192)	—	1215 (164)	1286	1310 (54)	1250 (182)	—	1223 (114)
10	—	—	1238 (57)	—	—	—	—	1218 (45)	—
10·5	1357 (181)	1309 (198)	—	1272 (178)	1307	1357 (71)	1306 (182)	—	1273 (110)
11	—	—	1279 (48)	—	—	—	—	1268 (46)	—
11·5	1395 (209)	1353 (211)	—	1308 (156)	1351	1412 (69)	1357 (185)	—	1318 (125)
12	—	—	1327 (56)	—	—	—	—	1303 (33)	—
12·5	1451 (189)	1397 (181)	—	1354 (131)	1399	1478 (65)	1408 (169)	—	1369 (127)
13	—	—	1352 (48)	—	—	—	—	1368 (29)	—
13·5	1506 (143)	1447 (162)	—	1397 (131)	1431	1521 (73)	1481 (180)	—	1411 (126)
14	—	—	1404 (53)	—	—	—	—	1409 (32)	—
14·5	1560 (158)	1466 (37)	—	1453 (117)	1488	1566 (62)	1505 (34)	—	1462 (105)
15·5	1624 (140)	—	—	1496 (76)	1542	1580 (46)	—	—	1475 (28)
16·5	1658 (117)	—	—	—	1616	—	—	—	—
17·5	1690 (70)	—	—	—	1669	—	—	—	—
18·5	1710 (40)	—	—	—	1684	—	—	—	—
19·5	1711 (23)	—	—	—	1668	—	—	—	—
Totals	1,740	1,496	562	1,374	*c.* 500	533	1,365	406	967

Table 5. *Average weights (in indoor clothing and without shoes) in kg. for pre-war series of German children*

Age (years): central values	Boys				Girls		
	Berlin			Hamburg	Berlin		
	'Higher schools' c. 1900	'Lower schools' c. 1900	Orphanage* 1919	'Higher schools' c. 1878	'Higher schools' c. 1900	'Lower schools' c. 1900	Orphanage* 1919
6	—	—	16·7 (86)	—	—	—	16·4 (74)
6·5	22·3 (45)	20·1 (128)	—	—	22·5 (14)	19·6 (110)	—
7	—	—	20·6 (82)	—	—	—	19·3 (45)
7·5	23·7 (101)	21·6 (189)	—	—	24·3 (42)	21·6 (159)	—
8	—	—	21·1 (58)	—	—	—	19·8 (55)
8·5	26·2 (156)	23·3 (198)	—	—	26·1 (37)	23·3 (164)	—
9	—	—	22·3 (74)	—	—	—	22·3 (47)
9·5	27·8 (168)	25·7 (192)	—	26·9	27·8 (54)	24·7 (182)	—
10	—	—	25·5 (57)	—	—	—	24·0 (45)
10·5	30·6 (181)	27·6 (198)	—	28·3	32·1 (71)	27·5 (182)	—
11	—	—	28·3 (48)	—	—	—	26·5 (46)
11·5	33·1 (209)	30·0 (211)	—	30·7	34·4 (69)	30·3 (185)	—
12	—	—	30·4 (56)	—	—	—	28·2 (33)
12·5	37·1 (189)	32·9 (181)	—	33·9	40·5 (65)	34·4 (169)	—
13	—	—	32·9 (48)	—	—	—	33·2 (29)
13·5	41·6 (143)	36·5 (162)	—	35·8	43·1 (73)	39·3 (180)	—
14	—	—	34·7 (53)	—	—	—	37·5 (32)
14·5	46·1 (158)	37·5 (37)	—	41·0	49·7 (62)	43·1 (34)	—
15·5	51·7 (140)	—	—	45·9	51·2 (46)	—	—
16·5	56·3 (117)	—	—	51·9	—	—	—
17·5	59·1 (70)	—	—	56·9	—	—	—
18·5	64·4 (40)	—	—	60·4	—	—	—
19·5	65·5 (23)	—	—	61·8	—	—	—
Totals	1,740	1,496	562	c. 500	533	1,365	406

* Nude weight given: allowances made for clothes as in table in § 3·5c.

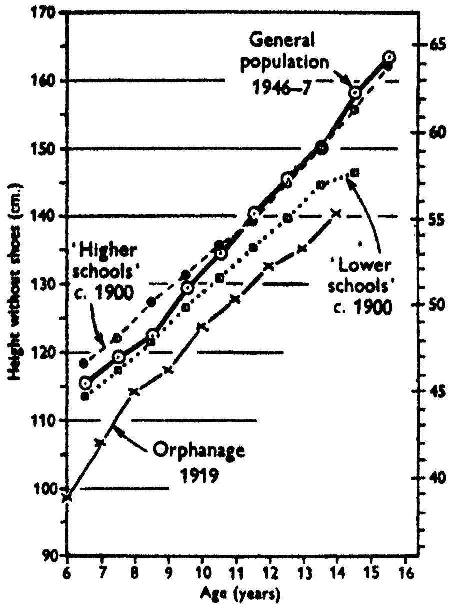


Fig. 9. Average heights for series of boys measured in Berlin. (Data for 1946–7 series in Table 3, and for other series in Table 4.)

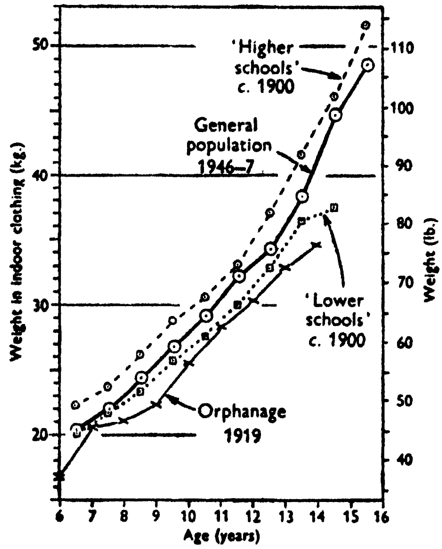


Fig. 10. Average weights (indoor clothing and without shoes) for series of boys measured in Berlin. (Data for 1946–7 series in Table 3, and for other series in Table 5.)

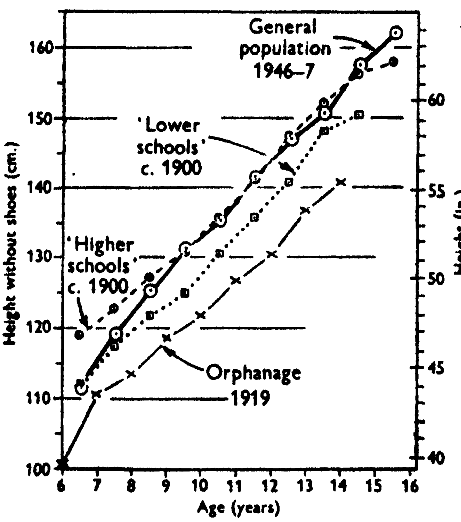


Fig. 11. Average heights for series of girls measured in Berlin. (Data for 1946–7 series in Table 3, and for other series in Table 4.)

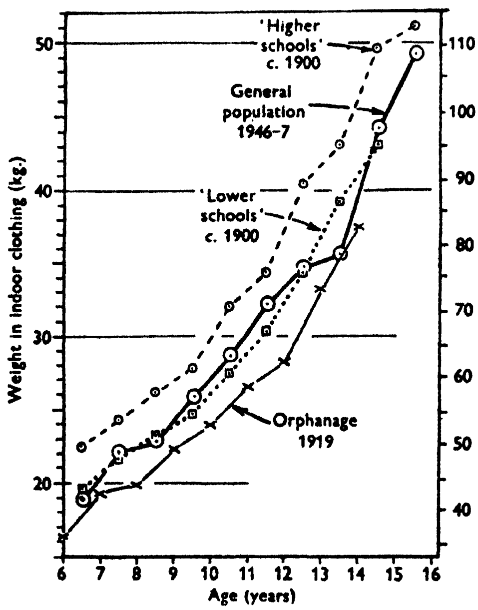


Fig. 12. Average weights (indoor clothing and without shoes) for series of girls measured in Berlin. (Data for 1946–7 series in Table 3, and for other series in Table 5.)

6. Comparisons of the post-war and earlier German and British series of children and adolescents

6·1. The mean heights and weights of recent British series given in Table 6 relate to:

(*a*) Boys' boarding schools assigned to the 'lowest economic standing' among boarding schools (group C), measurements being recorded in 1936–8 (Friend & Bransby, 1947).

(*b*) Male and female industrial workers included in a survey carried out by the Ministry of Food in 1943 (Kemsley, 1945).

(*c*) Male industrial workers included in a survey carried out by the Industrial Health Research Board of the Medical Research Council in 1929–32 (Cathcart *et al.* 1935).

Table 6. *Average heights and weights for series of British children and adolescents*

Age (years): central values	Height without shoes (mm.)			Weight in indoor clothing and without shoes (kg.)		
	Boys					
	Boarding schools 1936–8	Industrial workers 1943	Industrial workers 1929–32	Boarding schools 1936–8	Industrial workers 1943	Industrial workers 1929–32
11	1427 (284)	—	—	35·0	—	—
12	1443 (563)	—	—	36·8	—	—
13	1504 (888)	—	—	40·9	—	—
14	1565 (1537)	1558 (222)	—	45·9	45·9	—
15	1628 (1573)	1580 (806)	1540 (139)	51·7	49·1	44·9
16	1704 (544)	1637 (1310)	1603 (206)	58·9	54·0	50·7
17	1745 (412)	1673 (1560)	1656 (295)	63·6	57·7	55·8
18	1750 (152)	1684 (1240)	1686 (329)	65·1	59·9	59·2
19	—	1693 (729)	1691 (335)	—	61·1	60·6
	Girls					
	—	Industrial workers 1943	Industrial workers 1926	—	Industrial workers 1943	Industrial workers 1926
14	—	1545 (189)	—	—	46·5	—
15	—	1557 (698)	1557 (152)	—	49·5	45·8
16	—	1575 (1332)	1566 (213)	—	51·7	48·1
17	—	1579 (1622)	1570 (257)	—	52·8	49·5
18	—	1579 (1620)	1582 (259)	—	53·5	52·0
19	—	1582 (1377)	1585 (269)	—	54·0	52·7

(*d*) Female industrial workers included in a survey carried out by the same Board in 1926 (Cathcart *et al.* 1927).

6·2. The age 'curves' for these British and for the 1946–7 German (British zone except Berlin) series are shown in Figs. 13–16. For heights and weights of boys the German curves fall between those for the British pre-war boarding schools and industrial series, and the

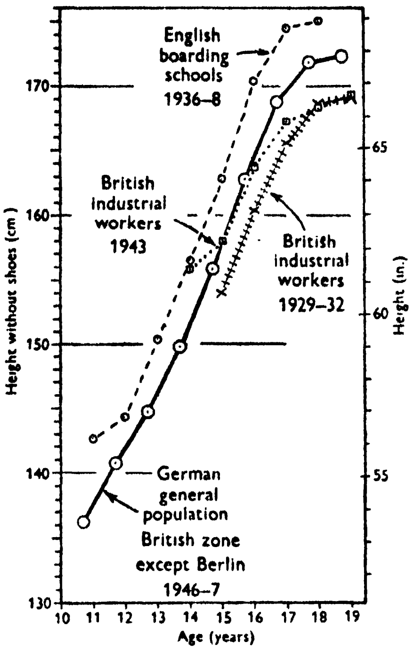


Fig. 13. Average heights for three British and a German series of boys. (Data for British series in Table 6, and for the German in Table 3.)

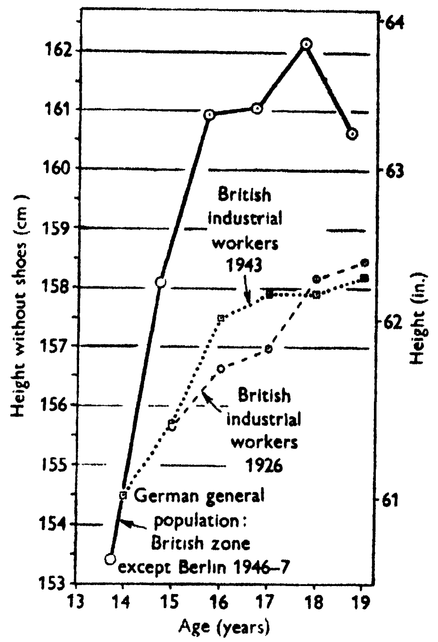


Fig. 14. Average heights for two British and a German series of girls. (Data for British series in Table 6, and for the German in Table 3.)

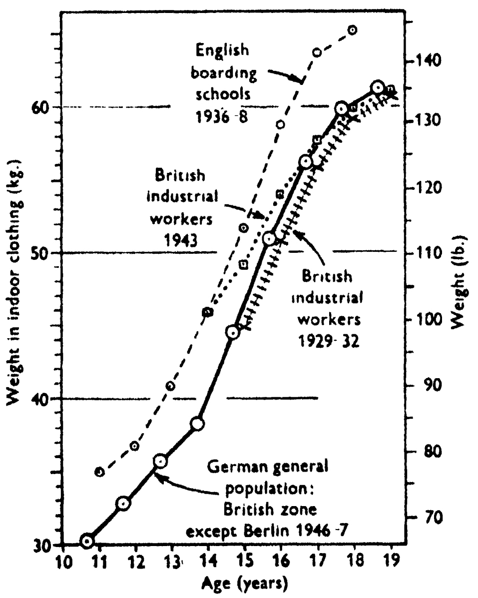


Fig. 15. Average weights (indoor clothing and without shoes) for three British and a German series of boys. (Data for British series in Table 6, and for the German in Table 3.)

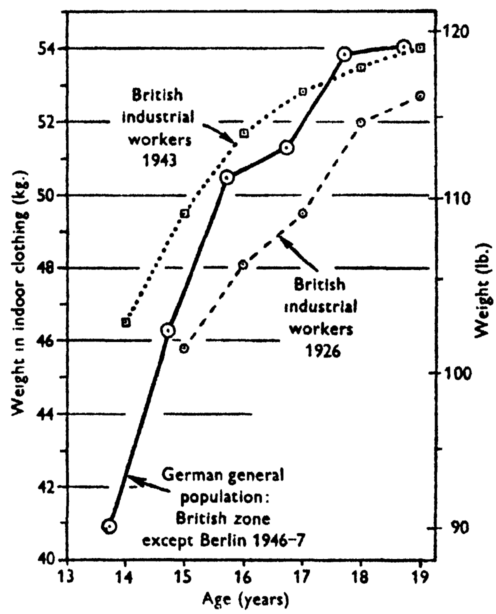


Fig. 16. Average weights (indoor clothing and without shoes) for two British and a German series of girls. (Data for British series in Table 6, and for the German in Table 3.)

three have very similar forms. It may be noted that the 1943 British series of male industrial workers differs appreciably from the 1929–32 representing the same class in showing, for both height and weight, greater means for ages under 18. For girls the German series is markedly superior for height, but inferior for weight, to the wartime British series.

6·3. Mean weights reduced to the same height at each age of the 1946–7 German series are given in Table 7 and plotted in Figs. 17 and 18 together with those for the other German and British series. For boys the various curves run nearly parallel courses and they suggest the following order:

English boarding schools 1936–8 (heaviest)	British industrial workers 1943	Berlin 'higher schools' *c.* 1900	British industrial workers 1929–32	Berlin orphanage 1919	Hamburg 'higher schools' *c.* 1878	German general population 1946–7	Berlin 'lower schools' *c.* 1900 (lightest)

Noteworthy features of this sequence are the high position of the British series of wartime industrial workers in it, and the fact that the post-war German series does not occupy an extremely low position. Its inferiority is most marked for the oldest age group (age 18). For ages 14–18 the mean reduced weights for the British wartime workers exceed the corresponding 1946–7 German values by amounts ranging from about 2–3 kg. (4·5–7·7 lb.).

The reduced weights for series of girls (Fig. 18) suggest the order:

Berlin 'higher schools' *c.* 1900 (heaviest)	British industrial workers 1943	Berlin 'lower schools' *c.* 1900	Berlin orphanage 1919	German general population 1946–7	British industrial workers 1926 (lightest)

A British pre-war standard is here below the post-war German, but the level of the British wartime series is decidedly high. The latter conclusion, which applies to both boys and girls, is satisfactory. It may be concluded that for the age range considered here—from the 14th to the 19th birthday—falls in the British 1943 standards for industrial workers less than 3 lb. would not be a serious matter. Reductions up to that limit, for both males and females, would mean that the class was still not inferior to its pre-war level.

7. Comparisons of average weights standardized to the same heights for the post-war German and Royal Air Force and other British and Dominion series of adults

7·1. The comparisons in this section are of average weights at different ages for various series of men and women. Nude weights are treated, allowances for clothes given in § 3·5 having been made where necessary, and all means are reduced to the height 1685 mm. for men and 1575 mm. for women by the method described in § 3·6. The series are:

(*a*) R.A.F. and Dominion series of aircrew. The survey of heights and weights was made in 1944. All the men were in an advanced stage of training (at O.T.U.'s) or on operational duties in Great Britain (unpublished).

(*b*) R.A.F. aircrew recruits. The majority of the men, measured in 1942, were direct entry civilians but some had served previously as R.A.F. ground staff. Many of the subjects of this survey must have been remeasured in the 1944 survey (*a*) above, though the exact proportion is unknown (Morant, 1943).

(*c*) R.A.F. aircrew in training. When measured in 1944 the men had just returned from training overseas (Canada and South Africa), (Morant & Gilson, 1945).

Table 7. *Average weights (in indoor clothing and without shoes) in kg. reduced at each age to the average height of the German (British zone except Berlin) 1946–7 series for various other German and British series of children and adolescents**

Age (years): central values	Boys							
	German					British		
	General population 1946–7	Berlin 'Higher schools' *c.* 1900	Berlin 'Lower schools' *c.* 1900	Berlin orphanage 1919	Hamburg 'Higher schools' *c.* 1878	Boarding schools 1936–8	Industrial workers 1943	Industrial workers 1929–32
6·5	21·2	22·2	21·3	22·7	—	—	—	—
7·5	23·1	23·7	23·0	24·2	—	—	—	—
8·5	25·4	26·1	25·0	25·2	—	—	—	—
9·5	27·5	27·7	27·2	27·4	27·7	—	—	—
10·5	30·0	30·7	29·4	30·6	30·2	—	—	—
11·5	32·2	33·3	31·9	33·2	32·7	34·5	—	—
12·5	35·0	36·6	34·8	36·1	35·7	37·4	—	—
13·5	38·0	40·8	38·6	39·3	38·7	41·2	—	—
14·5	43·4	45·5	42·1	42·2	44·4	46·2	46·5	—
15·5	49·9	51·5	—	—	50·6	52·5	52·2	50·7
16·5	55·4	57·7	—	—	55·9	58·4	57·7	56·4
17·5	59·1	60·4	—	—	59·6	62·4	60·9	60·0
18·5	61·1	65·7	—	—	63·4	—	63·2	62·6

Age (years): central values	Girls					
	German				British	
	General population 1946–7	Berlin 'Higher schools' *c.* 1900	Berlin 'Lower schools' *c.* 1900	Berlin orphanage 1919	Industrial workers 1943	Industrial workers 1926
6·5	20·3	22·0	20·9	20·7	—	—
7·5	22·2	23·8	22·6	21·9	—	—
8·5	24·2	25·8	24·5	23·9	—	—
9·5	26·4	27·5	26·3	26·2	—	—
10·5	29·0	31·9	29·0	29·0	—	—
11·5	32·0	34·3	32·4	32·3	—	—
12·5	35·9	40·1	37·3	36·9	—	—
13·5	40·0	43·0	41·3	42·1	—	—
14·5	45·4	49·9	46·5	47·1	49·0	—
15·5	49·9	52·8	—	—	53·0	49·5
16·5	51·3	—	—	—	54·1	51·1
17·5	53·5	—	—	—	53·9	52·7
18·5	54·0	—	—	—	55·4	53·8

* Another series of interest is of Belgian children of a working-class district of Brussels measured in the last three months of 1942, 1943 and 1944 (Ellis, 1945). To obtain large enough numbers at each age the data for the three years were pooled, giving n's ranging from 45 to 309, most representing more than 200 individuals. It is shown in the paper describing the records that the wartime children were *above* pre-war Belgian standards for height, and of the same order as the pre-war standards for weight. Reducing the mean weights (kg.) to the heights of the German 1946–7 series gives:

Age (years): central values	6·5	7·5	8·5	9·5	10·5	11·5	12·5	13·5	14·5	15·5
Boys	21·1	22·7	24·9	27·3	30·3	32·3	35·4	39·0	44·1	50·3
Girls	20·4	22·0	24·4	26·3	29·1	32·7	36·9	42·4	48·0	52·2

Compared with the series in the table above these reduced weights (in indoor clothing and without shoes) are extremely low for boys aged 6, 7 and 8, and low but not extreme in nearly all other cases

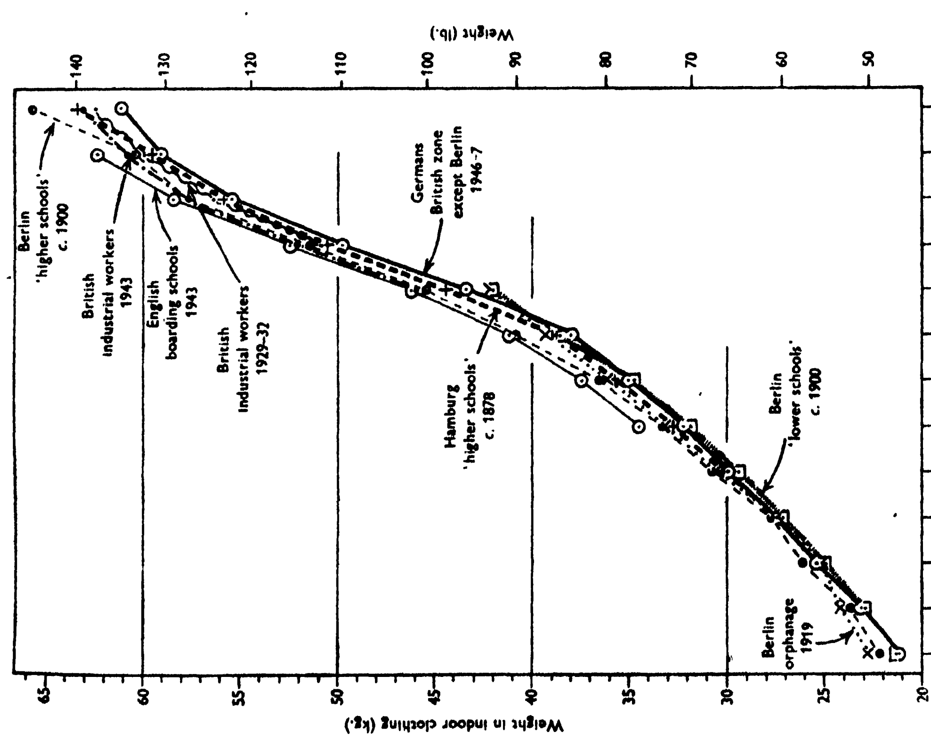


Fig. 17. Average weights (indoor clothing and without shoes) of German and British series of boys reduced to the average heights of the German 1946–7 series. (Data in Table 7.)

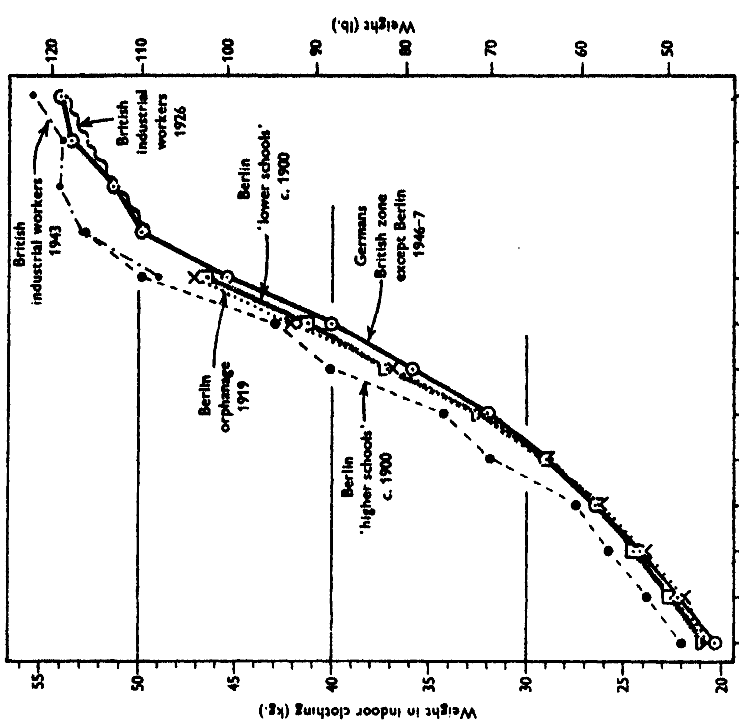


Fig. 18. Average weights (indoor clothing and without shoes) of German and British series of girls reduced to the average heights of the German 1946–7 series. (Data in Table 7.)

(*d*) R.N. pilots. The measurements used were those of the men when they were recruited for the Royal Navy from 1940 to 1946 (Morant, 1947).

(*e*) R.A.F. ground staff. The series represents a random sample of men recruited from 1940 to 1944, the measurements being those recorded at their first medical examinations (unpublished).

(*f*) Conscripts examined from 1 November 1917 to 31 October 1918. The data are for men examined by recruiting Medical Boards of the West Midland Region of England (Ministry of National Service, 1920). Those who were rejected for service in the Forces are included. At the time many who had previously been rejected were recalled for examination.

(*g*) Series of pre-war industrial workers. The surveys, carried out for the Industrial (Fatigue) Health Research Board of the Medical Research Council, were of women in 1926 (Cathcart *et al.* 1927) and of men in 1929–32 (Cathcart *et al.* 1935).

(*h*) Series of wartime industrial workers. The survey was carried out for the Ministry of Food in 1943 (Kemsley, 1945). Mean heights and weights given in the report were used to obtain the reduced weights shown in Figs. 20 and 21.

It should be noted that the Service series (*b*), (*d*), (*e*) and (*f*) are made up by men of whom all, or most, were civilians up to the time measurements were recorded.

7·2. The reduced weights for all the British and for the post-war German series are plotted in Figs. 19 and 20 (men) and 21 (women). In the case of the men there are far more series for ages up to 40 than for ages over 40. Considering the younger age range (17–40 years) only the series suggest the following order, though the sequence is not exactly the same for all ages within the range:

Operational aircrew: New Zealand — heaviest
Australian
Canadian
R.A.F.: Aircrew in training
Operational aircrew
R.N. pilot recruits (most civilian)
R.A.F. aircrew recruits (most civilian)
R.A.F. ground-staff recruits
Industrial workers, 1943
Employed industrial workers, 1929–32
Germans, 1946–7
Conscripts, 1917–18 and unemployed industrial workers, 1929–32 — lightest

This sequence is clearly significant. During the war, aircrew were the best-fed section of the British population, though the R.A.F. did not reach the high weight levels of the Dominion aircrew. The superiority of the New Zealand series is suggestive in view of the high nutritional and health standards of that country. The slight superiority of the aircrew in training over the British operational aircrew might be due to the fact that the former were measured immediately after their return from overseas; or the distinction might be partly due to the loss in weight, on the average, during the earlier stages of wartime operational tours, which was found to be of the order 1·5 lb. (Reid, 1947). Aircrew recruits (R.N. pilots and R.A.F.) come next, followed by R.A.F. ground-staff recruits. When plotted together, the R.A.F. ground-staff recruits (Fig. 19) are seen to have greater mean

weights than the wartime and pre-war industrial workers (Fig. 20) for ages up to about 28 years, but smaller means for all later ages represented. The younger men accepted for service, chiefly on medical grounds, must have been heavier, on the average, than the

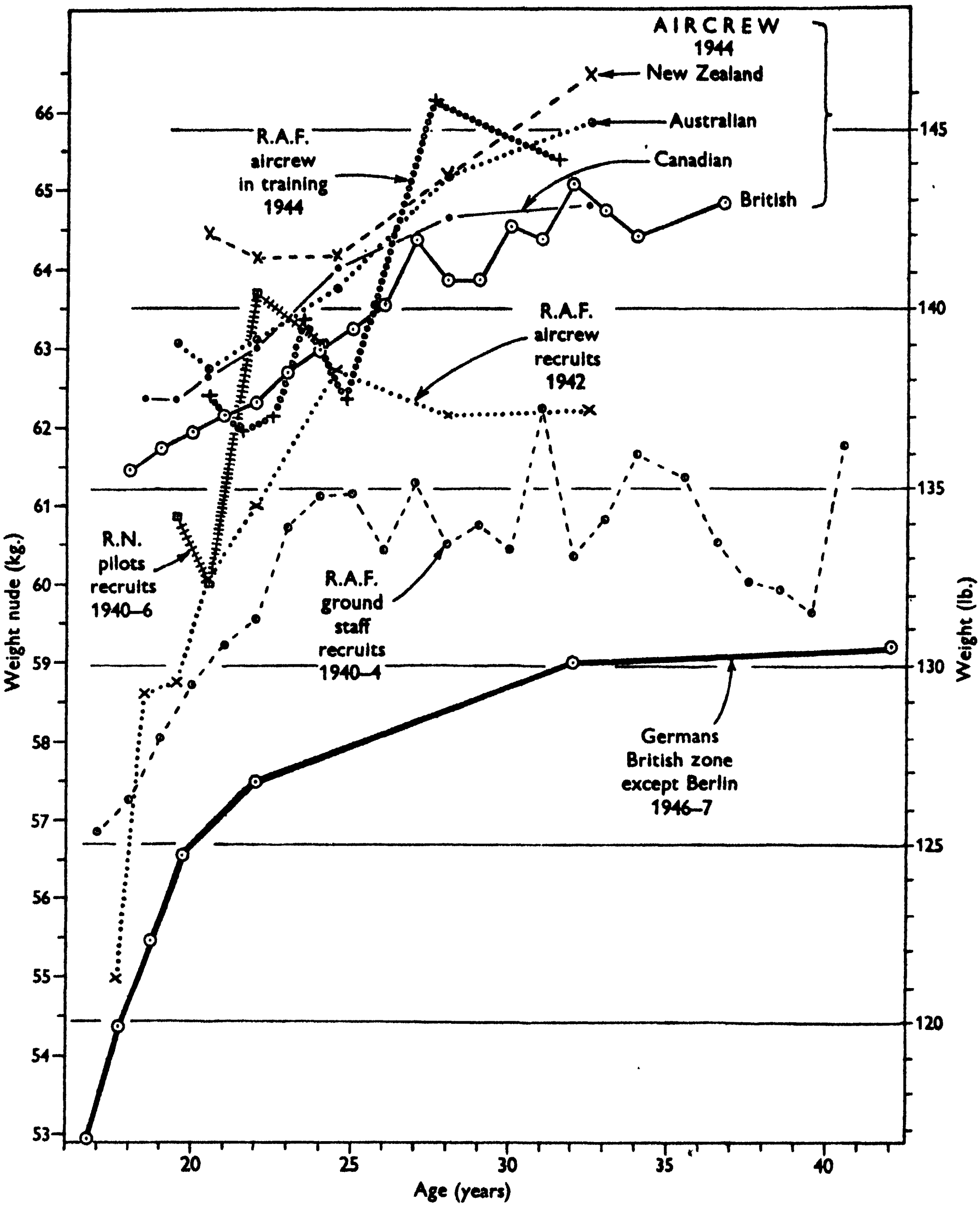


Fig. 19. Average weights (nude) standardized to height 1685 mm. for the post-war German series of civilians and series of R.N. and R.A.F. personnel. The smallest series are of R.N. pilots ($n = 200$), R.A.F. aircrew in training (529) and New Zealand aircrew (550), all the others being of 1600 or more men.

population from which they were drawn, but for the older men the position was reversed. For men up to age 40 it is surprising to find that the post-war German series has to be classed rather above two earlier British ones. Compared with them it shows superiority in average weights for ages up to about 25 years, and the same level as theirs for ages 25–40.

For ages over 40 (Fig. 20) the situation is markedly different. The German series clearly falls to the lowest place and the wartime industrial workers fall below the pre-war employed workers of the same class, though remaining above the low levels of the pre-war unemployed

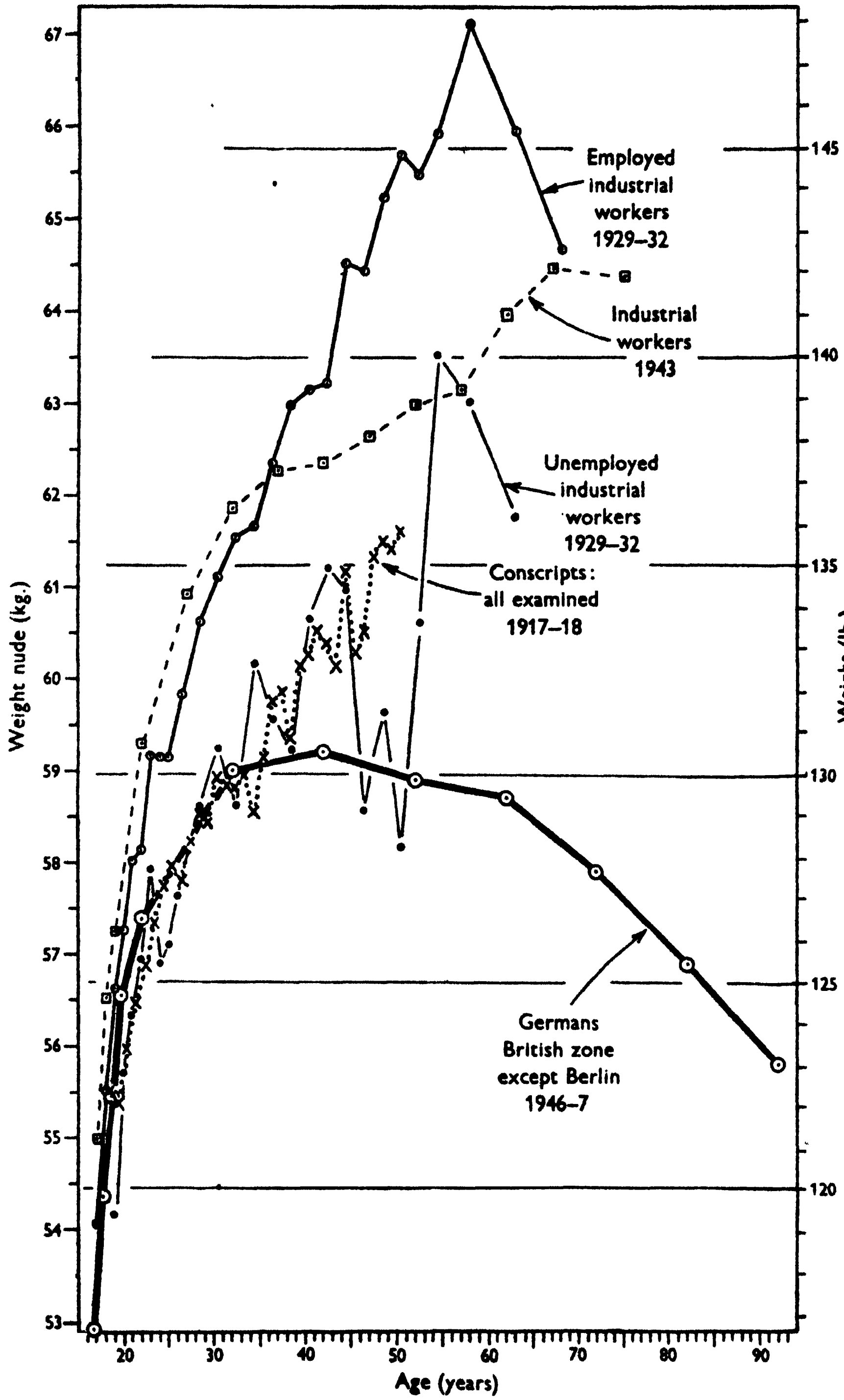


Fig. 20. Average weights (nude) standardized to height 1685 mm. for the post-war German and British series of civilian men. The smallest series is of 1,328 unemployed industrial workers and each of the other four relates to more than 10,000 men.

industrial workers and the 1917–18 conscripts. Food rationing affects most markedly the weights of the middle aged and aged.

This is shown again by the few female series (Fig. 21). The maximum of the age curve for the German women is in the twenties—compared with the forties for the men (Fig. 20)—and at age 57 the German mean is nearly 10 kg. (22 lb.) below that of the British wartime workers. The corresponding series of men show a maximum divergence at age 75 of nearly 7 kg. (15 lb.). In general the absolute, and still more clearly the relative, weight losses of German adults since the war must have been greater for women than for men.

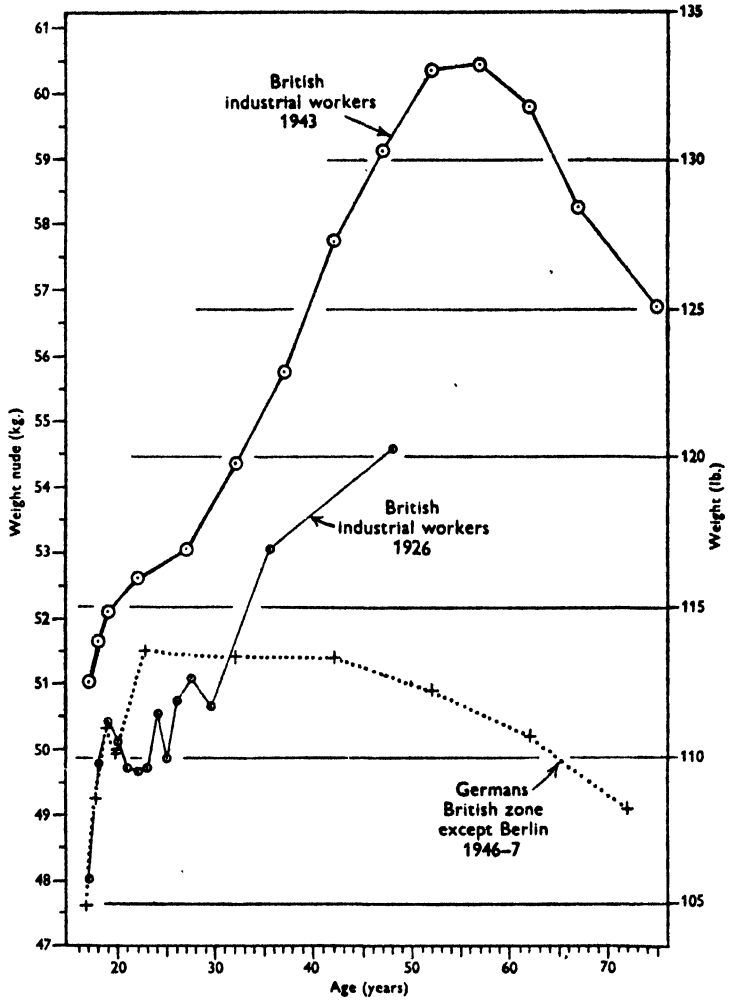


Fig. 21. Average weights standardized to height 1575 mm. for two British and a German series of women. The 1926 series is of 3,000, and the 1943 series of 31,200, British women.

7·3. The normal form of the age curve for weight in the case of both males and females shows continuous increase of the average up to some age about 60 years and then a decline in old age. Of the series considered the only ones which fail to conform to this pattern are the R.A.F. aircrew and ground-staff recruits (Fig. 19), which show increase to age 25 and then remain at about the same level, the German series of men (Fig. 20, maximum about 40 years) and women (Fig. 21, decline from ages 19–20 and maximum in early twenties),

and the pre-war British series of women industrial workers (Fig. 21, abnormal decline between ages 19 and 24). All except the last of these series had been subject to civilian rationing, but the 1943 men and women industrial workers, who show the normal form of curve, were also subject to it. The peculiarities of the recent series are probably due to the interaction of rationing, of medical selection in some cases, and of other imponderable factors.

7·4. The data presented provide standards which may be of use in future comparisons. It might be asked, for example, what losses in weight for a particular civilian or Service population for which repeated records are available should be considered a matter of serious concern. It is clear that such a question should only be considered with reference to particular years of age or age groups covering a few years. Statements regarding all men, say, or all boys, would be of no scientific value.

7·5. Records of average weights appreciably lower than those for the post-war Germans are not easy to find. Comparison is made below with means for broad age groups given for United States prisoners in Santo Tomas camp, Manila, Phillipine Islands, 36 months after internment (Brown, 1946). The corresponding means for the Germans can only be estimated approximately as the data for them are in decennial age groups and no finer age distribution for the American series is given. Also the German weights are standardized to heights which are probably rather below the unknown averages for the American series:

Mean weights (nude) in lb.

Ages		American	German	Difference
Men	19–40	124	128	4
	41–60	122	130	8
	Over 60	119	128	9
Women	19–40	101	113	12
	41–60	100	112	12
	Over 60	96	109	13

The differences here are substantial and greater for women than for men. The evidence considered in this paper is consistent in showing for all adult ages that loss of weight due to restricted rations is greater for females than for males.

This paper is chiefly concerned with a treatment of unpublished official records. The writer acknowledges with thanks permission to use such material given by: (*a*) the Public Health Branch of the Control Commission for Germany (British Element), by courtesy of Brigadier W. Strelley Martin, M.C., Public Health Adviser, Health Branch, C.C.G., and F. D. G. Bailey, Esq., formerly Nutrition Officer, Health Branch C.C.G.; (*b*) the Medical Directorate of the Air Ministry, by courtesy of Air Vice-Marshal P. C. Livingston, C.B., C.B.E., A.F.C., F.R.C.S., Director-General of Medical Services, Royal Air Force; (*c*) the Ministry of Food, by courtesy of the Chief Scientific Officer to the Ministry.

REFERENCES

BROWN, B. H. (1946). A detailed report on the weights and weight losses of twenty-four men in Santo Tomas internment camp. *J. Lab. Clin. Med.* **31**, 1129–32.

CATHCART, E. P., HUGHES, D. E. R. & CHALMERS, J. G. (1935). The physique of man in industry. *M.R.C., Industr. Heal. Res. Board, Report* no. 71.

CATHCART, E. P., BEDALE, E. M., BLAIR, C., MACLEOD, K. & WEATHERHEAD, M. (1927). The physique of women in industry. *M.R.C., Industr. Fat. Res. Board, Report* no. 44.

CONTROL COMMISSION FOR GERMANY (BRITISH ELEMENT), PUBLIC HEALTH BRANCH, NO. 1 NUTRITION SURVEY TEAM (1947). Report on a body-weight survey in the British zone.

COON, C. S. (1939). *The Races of Europe*. New York: Macmillan.

DAVIDSOHN, H. (1919). Die Wirkung der Aushungerung Deutschlands auf die Berliner Kinder mit besonderer Berücksichtigung der Waisenkinder der Stadt Berlin. *Z. Kinderheilk.* **21**, 349–407.

DUNSTAN, W. R. (1925). Height and weight of school children in an English rural area. *Metron*, **5**.

ELDERTON, E. M. (1914). Height and weight of school children in Glasgow. *Biometrika*, **10**, 288–339.

ELLIS, R. W. B. (1945). Growth and health of Belgian children during and after the German occupation (1940–1944). *Arch. Dis. Childhood*, **20**, 97–109.

FRIEND, G. E. (1935). *The Schoolboy: a Study of his Nutrition, Physical Development and Health*. Cambridge: Heffer.

FRIEND, G. E. & BRANSBY, E. R. (1947). Physique and growth of schoolboys. *Lancet*, 8 Nov., p. 677.

KEMSLEY, W. F. F. (1945). *A Preliminary Account of the Ministry of Food's Survey into the Weight and Height of the Population during* 1943. Ministry of Food Surveys Branch, DPD/W/6.

KEMSLEY, W. F. F. (1946). *Body Weight Survey: Report on the Weighings taken between April* 1943 *and October* 1945. Ministry of Food Surveys Branch, DPD/W/7.

KEMSLEY, W. F. F. (1947). *Body Weight Survey: Variations in Body Weight*—1943 *to* 1946. Ministry of Food Surveys Branch, DPD/W/8.

KOTELMANN, K. (1879). Die Körperverhältnisse der Gelehrtenschüler des Johanneums in Hamburg. *Z. kgl. preuss. stat. Bur., Berlin*, **19**, 1–16.

MINISTRY OF NATIONAL SERVICE, 1917–19 (1920). *Report, Vol.* 1, *upon the Physical Examination of Men of Military Age by National Service Medical Boards from November* 1*st*, 1917 *to October* 31*st*, 1918. H.M.S.O.

MORANT, G. M. (1943). Preliminary report on body measurements of 2,400 candidates for air-crew. *Flying Personnel Res. Comm.* no. 538.

MORANT, G. M. (1947). The heights and weights of Royal Naval pilots compared with those of R.A.F. pilots. *Flying Personnel Res. Comm.* no. 673.

MORANT, G. M. & GILSON, J. C. (1945). A report on a survey of body and clothing measurements of Royal Air Force personnel. *Flying Personnel Res. Comm.* no. 633*a*.

MUMFORD, A. A. (1927). *Healthy Growth: a Study of the Relation between the Mental and Physical Development of Adolescent Boys in a Public Day School*. Oxford University Press.

RANKE, O. (1905). Beiträge zur Frage kindlichen Wachstums: anthropologische Untersuchungen ausgeführt an holsteinischen Kindern von der Geburt bis zum vollendeter 15 Jahre. *Arch. Anthrop.*, N.F., **3**, 161–80.

REID, D. D. (1947). Some measures of the effect of operational stress on bomber crews (F.P.R.C. 605). *Psychological Disorders in Flying Personnel of the Royal Air Force investigated during the War* 1939–1945, Air Publ. 3139, pp. 245–58. H.M.S.O.

RIETZ, E. (1903). Das Wachstum der Berliner Schulkinder während der Schuljahre. *Arch. Anthrop.*, N.F., **1**, 30–40.

SCHLESINGER, E. (1917). Das Wachstum der Knaben und Jünglinge vom 6 bis 20 Lebensjahr. *Z. Kinderheilk.* **16**, 265–304.

SCHMID-MONNARD, K. (1895). Über den Einfluss der Jahreszeit und der Schule auf das Wachstum der Kinder. *Jahr. Kinderheilk.* **40**, 84–107.

TESTING THE SIGNIFICANCE OF CORRELATION BETWEEN TIME SERIES*

By G. H. ORCUTT and S. F. JAMES, *Department of Applied Economics, University of Cambridge*

I. Introduction

Verification that two things are related to each other is achieved by showing empirically that there is a correspondence between their behaviours greater than could be expected by chance. Where continued experiment is not possible the data available are usually severely limited in both quantity and range of variation and it therefore becomes essential to have precise ideas whether the agreement with a hypothesis is actually a verification of it or whether it is more reasonable to consider the agreement as merely the result of a chance correspondence. This paper discusses the problem of deciding when a correlation between two economic time series is great enough to make it unreasonable to assume that the series are unrelated.

The testing of the significance of a correlation involves a comparison with what would have been obtained between non-related series thought to be analogous to the observed series. And, of course, the significance found for the correlation will depend upon the analogy deemed to be appropriate. The choice of an analogy depends upon experience as to which aspects of the real series being correlated are vital, in the sense that they affect the probability of obtaining chance correlations between non-related series. We can never be certain that some important aspect has not been overlooked, but, as our experience is broadened and we learn to take into account more and more factors, the chances of our running into a situation in which the analogy we choose is actually misleading becomes less and less. If we were forced to base our choice of an appropriate analogy to use in any individual situation on the data of that situation alone, the uncertainty of our tests of significance would be very great. Usually, however, the situation is one which we have learned to be similar to some larger class of experiences, and it is this larger class of experiences that furnishes a greater measure of information as to the analogy appropriate to all members of the class.

The most commonly used sampling model for generating series of independent terms is to draw them at random from a normal population of values; and in applying tests based upon this sampling model one must take account of the length of the series being dealt with. Fortunately, there is some evidence that tests of significance based on this sampling model are insensitive to variation of the frequency distribution of the population of values from which the random sampling is done (Pearson, 1931), and this makes it reasonable to apply such tests even when little is known of the frequency distributions from which the items of our real series have been drawn. There is, however, one obvious point at which the analogy underlying such tests of significance may break down when one is concerned with economic time series, namely, if the consecutive terms are really correlated. In economic time series, in meteorological time or spatial series, or, for that matter, in biological time series, autocorrelation usually exists. Production, or employment, or price-level series never go directly from

* Mr James was largely responsible for the part embodied in section III of this study. We both wish to express our appreciation for the large measure of assistance given us by Mr Richard Stone.

high values to low values, but, instead, high values are followed by values which are also high and a transition from high values to low values only takes place over a period of time. How closely successive values are related will of course partly depend upon the time between measurements and, as this time is made shorter and shorter, successive values of the series of measurements become more and more like their immediate neighbours. Autocorrelations are often very high and remain high even as the series are lengthened, whereas in random series the autocorrelations are small tending to zero as the series increase in length. In economics, most of the material that we wish to investigate for relationships exhibits autocorrelation, and there is a real need for a test of significance for correlations which is based on a more realistic sampling model.

Bartlett (1935) obtained the following large sample approximation of the variance of the sample correlations between two autocorrelated series having a true correlation of zero:

$$\operatorname{var} r \sim \frac{n+2[(n-1)\rho_1\rho_1' + (n-2)\rho_1^2\rho_1'^2 + \ldots + \rho_1^{n-1}\rho_1'^{n-1}]}{n^2}, \tag{1}$$

or more approximately
$$\operatorname{var} r \sim \frac{1}{n}\frac{1+\rho_1\rho_1'}{1-\rho_1\rho_1'}, \tag{2}$$

where ρ_1 is the true value of the first autocorrelation of one of the two series and ρ_1' is that of the other. This is based on the assumption that each of the series was generated by the following type of process

$$x_t = \rho_1 x_{t-1} + \epsilon_t, \tag{3}$$

where the random error term, ϵ_t, is independent of x_{t-1} and $E(\epsilon_t) = 0$. Since then, Bartlett (1946, 1947) and Quenouille (1947) have given a large sample approximation of var r for correlations between any two autoregressive schemes having a true correlation of zero.* For linear autoregressive schemes it is

$$\operatorname{var} r \sim \frac{n+2[(n-1)\rho_1\rho_1' + (n-2)\rho_2\rho_2' + \ldots + \rho_{n-1}\rho_{n-1}']}{n^2}, \tag{4}$$

or more approximately
$$\operatorname{var} r \sim \sum_{t=-\infty}^{\infty} \rho_t\rho_t'/n. \tag{5}$$

Quenouille (1947) has given a convenient method of evaluating an expression such as (5).

Now while these formulae make it perfectly clear that, in interpreting the significance of a correlation between two series, it is necessary to take account of their autocorrelations, they have, as is recognized, certain practical limitations. Besides being based on the true autocorrelations, which are never obtainable in practice, they involve large sample assumptions which might not be reasonable for series of twenty or thirty items with which the economist must usually deal. It is also evident that, given only the formulae for var r, it is impossible, even neglecting the above considerations, to apply a test of significance without some knowledge concerning the shape of the distribution of sample correlations.

With the above difficulties in mind, we decided that a sampling experiment would give some guidance as to a reasonable procedure for carrying out tests of significance in the case of small samples. Since we could only carry out a rather limited sampling experiment, we were anxious that the sampling model used should generate unrelated series which were as analogous as possible to economic time series. In this way we might hope to obtain the

* The reader is also referred to a useful paper by Moran (1947), which gives the formula for the variance of the covariance between two series having known autoregressive properties.

maximum guidance in a region of practical interest. Yule (1921, 1926, 1927), Wold (1938) and Kendall (1944, 1945, 1946) have stressed that for most economic time series an autoregressive scheme is probably more relevant than the assumption of exact harmonic oscillation and each of the above has made considerable use of the linear second-order autoregressive scheme in studies of economic time series. Orcutt (1948*a*) tested the hypothesis that the economic time series used by Tinbergen (1939) might be considered to have been obtained by drawings from a single population of linear stochastic series all having the same underlying autoregressive structure. This hypothesis was brought to a test by comparing the means and variances at each lag of the correlograms of the economic series with the means and variances at corresponding lags of the correlograms of several sets of other series constructed according to a variety of models. On the basis of these comparisons and also a similar set of comparisons of correlograms of first differences, the conclusion was reached that so far as the evidence went the set of 52 economic series might have been obtained by drawings from the population of series generated by the model

$$Y_{(t+1)} = Y_t + 0{\cdot}3(Y_t - Y_{(t-1)}) + \epsilon_{(t+1)}, \tag{6}$$

where ϵ is random in time and has an expected value of zero.

Since equation (6) appears to us as the best available model for generating non-related series which are analogous to economic time series, we have used it as a basis for generating the series used in the remainder of this paper. It should be noted that equation (6) does not generate stationary series but rather a Brownian type of movement having no true mean. See Wold (1938, p. 53) for the distinction between stationary and evolutive time series. On the other hand, the series generated by equation (6) are not explosive in the sense that they tend to deviate from any given point or set-up oscillations of ever increasing amplitude. Six 30-item segments which were selected without regard to shape from a long series generated by equation (6) are shown in Fig. 1. Since the formulae given earlier for var r were derived on the assumption of stationary autoregressive processes, it is clear that on this account alone it would not be safe without additional evidence to apply them to correlations between non-stationary series such as generated by equation (6).

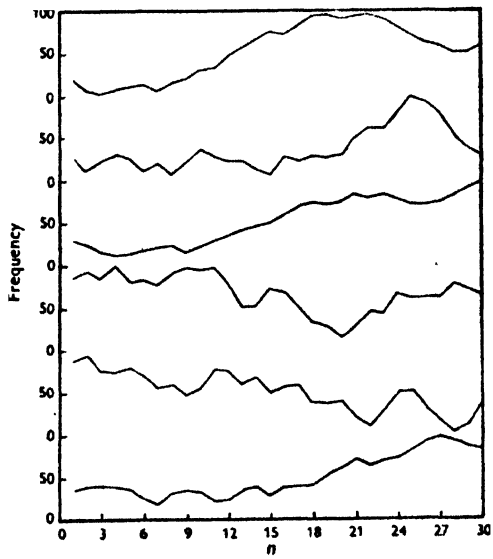


Fig. 1. 30-item segments of constructed series.

II. Empirical distributions of correlations between non-related series generated by the autogressive process $Y_{(t+1)} = Y_t + 0{\cdot}3(Y_t - Y_{(t-1)}) + \epsilon_{(t+1)}$

We therefore sought to obtain empirically some idea of the distribution of correlations to be expected by chance between series drawn at random from the population of non-related series generated by the stochastic process of equation (6). In order to do this we first constructed a series of 3240 items by means of (6). The random elements used were two digit

numbers taken from Tables of Random Sampling Numbers by M. G. Kendall and B. Babington Smith (1939). They were read from left to right and double zeros were omitted. To obtain a true mean of zero, 50 was subtracted from each. Thus the random numbers used were drawn from a population having a rectangular distribution and a range of -49 through 0 to $+49$.

Our long series of 3240 items was first divided into 36 series each of 90 items. Each of these 36 series was then divided into three series of 30 items. The first 30-item series of the first 90-item series was labelled 1 A, the second 1 B and the third 1 C. The first 30-item series of the second 90-item series was labelled 2 A, the second 2 B and so on.

By the use of the usual product-moment formula for the correlation coefficient, correlations were obtained between all possible pairs of A series, between all possible pairs of B series and between all possible pairs of C series. That is, series 1 A was correlated in turn with series 2 A, 3 A, ..., 36 A; series 2 A was correlated in turn with series 3 A, 4 A, ..., 36 A, and so on. Thus we obtained 630 correlations between pairs of A series, 630 between pairs of B series and 630 between pairs of C series, making a total of 1890 correlations between different pairs of our 108 series each of 30 items. Then labelling the first 60 items of each 90-item series, series (A + B), we found the correlations between all possible pairs of the 36 (A + B) series. Having obtained these 630 correlations between series of 60 items we then found the 630 correlations obtained by correlating all the pairs of our 36 series of 90 items each. These series are labelled the (A + B + C) series. The labour of obtaining the above correlations was rather large but the calculations were considerably facilitated by use of a new type of calculating machine (Orcutt, 1948*b*). Now while it can be shown that, when the series are independent, the above sampling procedure will lead to unbiassed estimates of the moments of the population of correlations between series drawn at random from the universe of series generated by the autoregressive process used, it is evident that the 1890 correlations obtained are not completely independent, so that, while the effective number is substantially greater than 108 (the number of independent series), it, nevertheless, is considerably less than 1890. The reason for using each series a large number of times is simply the saving of labour.

Table 1 gives the frequency distribution for our constructed series with $n = 30$, $n = 60$, and $n = 90$, of the ratio of the mean square successive difference to the variance. This ratio is usually denoted by δ^2/s^2 and for infinite series $\delta^2/s^2 = 2(1-\rho_1)$. For a discussion of this ratio and tabulations of its probability distribution for random series, see von Neumann (1941, 1942) and Hart & von Neumann (1942).

Table 2 gives the frequency distributions of the correlations obtained between pairs of each of the sets of 30-item series and their total together with the frequency distributions of correlations obtained between pairs of the 60-item series and the 90-item series. Fig. 2 shows graphically the frequency distribution of the correlations for $n = 30$ together with a curve showing the frequency distribution of correlations between pairs of non-related random normal series with $n = 5$.* We tested the fit of this theoretical curve by means of a χ^2 test with 20 classes and 19 d.f., since the variance of the theoretical curve has been approximately fitted, and obtained a probability of less than 0·01. Figs. 3 and 4 show the distributions of the correlation coefficient which we obtained with $n = 60$ and $n = 90$, respectively. On the first we have drawn the curve for random series with $n = 5$ and on the second the

* The value 5 was chosen to make the variance $1/(n-1)$ agree as closely as possible with the observed variance 0·2736. Similarly $n=6$ gives close agreement with the observed 0·2061 in Fig. 4.

curve for random series with $n = 6$. With 17 d.f., we obtained a χ^2 of 16·3 corresponding to a probability of about 0·5 in the first case and with the same degrees of freedom we obtained a χ^2 of 15·9 corresponding to a probability of about 0·7 in the second case. Table 3 gives the cumulative frequency distributions from the total of the 30-item series, and

Table 1. *Frequency distribution of δ^2/s^2 for series generated by*
$$Y_{(t+1)} = Y_t + 0{\cdot}3(Y_t - Y_{(t-1)}) + \epsilon_{(t+1)}$$

δ^2/s^2	Frequency		
	$n=30$	$n=60$	$n=90$
0·00–0·10	40	23	30
0·10–0·20	27	9	6
0·20–0·30	23	1	0
0·30–0·40	5	1	0
0·40–0·50	4	1	0
0·50–0·60	4	1	0
0·60–0·70	2	0	0
0·70–0·80	2	0	0
0·80–0·90	1	0	0
Total	108	36	36
Mean δ^2/s^2	0·195	0·104	0·062

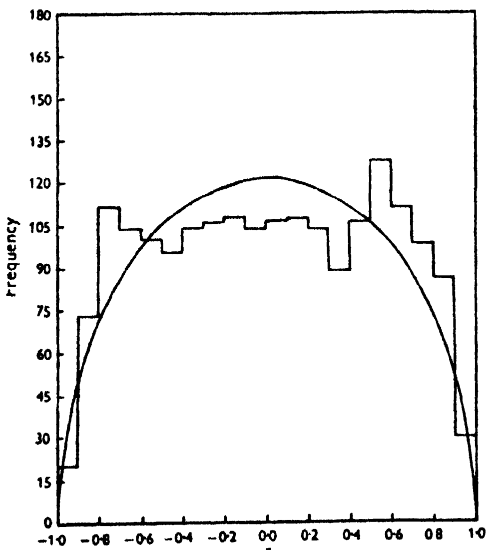


Fig. 2. Frequency distributions of r, $n = 30$.

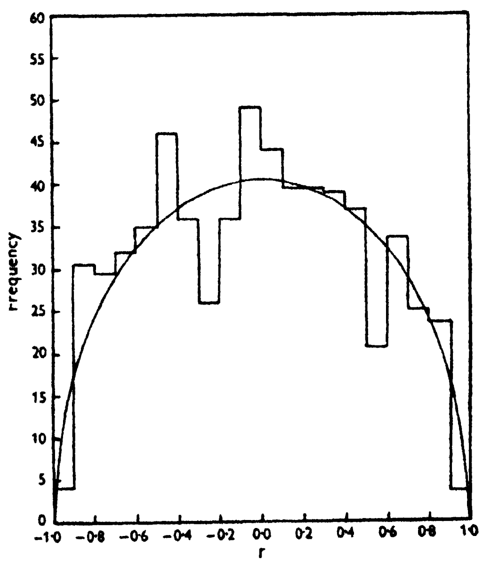


Fig. 3. Frequency distribution of r, $n = 60$.

for the 60- and 90-item series. The above application of the χ^2 test for testing the adequacy of the theoretical curves is admittedly rough, both because we have only approximately fitted the variances of the theoretical curves to our empirical distributions by our choice of n for the distributions of correlations between random series, and because the correlations making up our empirical distributions are not completely independent. The effect of both the above

Table 2. *Frequency distributions of r*

r	Frequency*					
	$n=30$				$n=60$ Series A+B	$n=90$ Series A+B+C
	Series A	Series B	Series C	Total A, B, C		
−1·0 to −0·9	6·5	12·0	2·0	20·5	4·0	3·5
−0·9 −0·8	28·0	32·5	12·0	72·5	30·5	10·5
−0·8 −0·7	35·0	49·0	27·5	111·5	29·5	25·0
−0·7 −0·6	37·0	37·5	29·0	103·5	32·0	31·0
−0·6 −0·5	33·0	31·0	36·0	100·0	35·0	37·0
−0·5 −0·4	37·5	25·5	32·0	95·0	46·0	40·0
−0·4 −0·3	43·5	18·5	41·5	103·5	36·0	38·5
−0·3 −0·2	36·0	32·0	38·0	106·0	26·5	37·0
−0·2 −0·1	30·0	31·5	47·0	108·5	36·0	44·5
−0·1 −0·0	34·0	47·0	22·5	103·5	49·0	45·5
+0·0 +0·1	31·0	39·0	36·5	106·5	44·0	54·0
+0·1 +0·2	30·0	31·5	46·0	107·5	39·5	46·0
+0·2 +0·3	34·5	26·5	42·5	103·5	39·5	30·5
+0·3 +0·4	29·5	25·0	34·0	88·5	39·0	46·5
+0·4 +0·5	30·0	28·0	48·0	106·0	37·0	38·5
+0·5 +0·6	40·5	44·5	43·5	128·5	21·0	31·5
+0·6 +0·7	36·5	39·0	35·5	111·0	33·5	27·5
+0·7 +0·8	37·0	30·0	31·0	98·0	25·0	30·0
+0·8 +0·9	25·0	37·0	24·0	86·0	23·5	11·5
+0·9 +1·0	15·5	13·0	1·5	30·0	3·5	1·5
Total	630·0	630·0	630·0	1890·0	630·0	630·0
Mean	0·0048	0·0004	0·0439	0·0164	−0·0248	−0·0023
$s^2=\Sigma r^2/n$	0·2844	0·3071	0·2293	0·2736	0·2446	0·2061
s	0·5333	0·5542	0·4788	0·5231	0·4945	0·4540
β_1	0·0047	0·0002	0·0335	0·0044	0·0140	0·0002
β_2	1·8125	1·7726	1·9238	1·8476	1·9422	2·0357

Table 3. *Cumulative frequency distributions with positive and negative r's combined*

$\|r\|$	Fraction greater than $\|r\|$		
	$n=30$	$n=60$	$n=90$
0·00	1·00	1·00	1·00
0·10	0·89	0·85	0·84
0·20	0·77	0·73	0·70
0·30	0·66	0·63	0·59
0·40	0·56	0·51	0·46
0·50	0·46	0·38	0·33
0·60	0·33	0·29	0·22
0·70	0·22	0·18	0·13
0·80	0·11	0·10	0·04
0·85	0·06	0·05	0·03
0·90	0·03	0·01	0·01
0·95	0·01	0·00	0·00
1·00	0·00	0·00	0·00

* Values of r were calculated to two places of decimals. Those on the border of two class-intervals were allocated one-half to each interval. The same applies to Tables 4 and 5.

circumstances will be to tend to exaggerate the χ^2's obtained and so underestimate the true probabilities that discrepancies between the empirical frequency distributions and the theoretical distributions are chance results.

Table 4 gives the distribution of 306 first-order partial correlations where $n = 30$ and two explanatory series are used. These 306 comprise the combinations available from our correlations between 30-item series using only partial correlations of the types, $r_{i(i+1).(i+2)}$, $r_{i(i+2).(i+1)}$ and $r_{(i+1)(i+2).i}$. Their frequency distribution is shown on Fig. 5 along with a curve showing the frequency distribution of correlations between pairs of non-related random series with $n = 5$.

Table 5 gives the distribution of 306 multiple correlations involving two explanatory series. These correlations were obtained from the above partial correlations and zero-order correlations. They are of the types, $R_{i.(i+1)(i+2)}$, $R_{(i+1).i(i+2)}$ and $R_{(i+2).i(i+1)}$. Their frequency distribution is shown on Fig. 6.

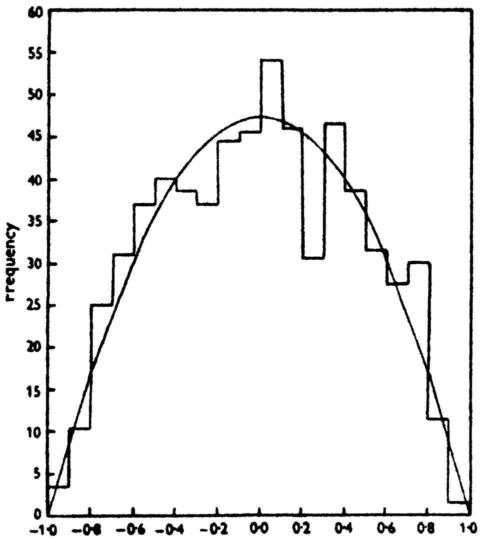


Fig. 4. Frequency distribution of r, $n = 90$.

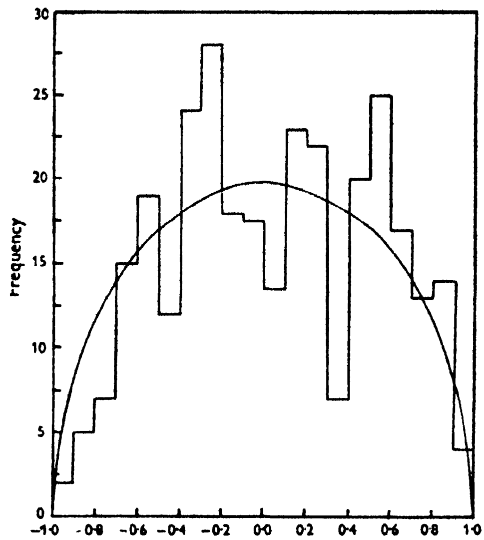


Fig. 5. Frequency distribution of first order partial correlations, $n = 30$.

In the next section we shall investigate whether sample-values can be used for unknown parameters in determining probabilities for the testing of correlations, but there are certain observations which might usefully be made at this stage. In the first place, it should be evident that the empirical distributions of this section provide a test of the null hypothesis that a given sample correlation might have occurred between series drawn at random from the population of series generated by equation (6). Since the distributions change very slowly with n, the length of the series, it should be possible to interpolate for a particular value of n with as much accuracy as our distributions justify. The position with regard to partial and multiple correlations is not so satisfactory since we have merely obtained, for $n = 30$, distributions of first-order partial correlations and multiple correlations involving two explanatory series. It is hoped, however, that these will be sufficient to give some idea of how high these coefficients must be in order to provide significant evidence against the null hypothesis and to throw some light on the possibility of getting very high multiple correlations between non-related series when n is less than 30 and four or five explanatory series are used.

Table 4. *Frequency distribution of first order partial correlations between series with $n = 30$*

$r_{ij.k}$	Frequency
−1·0 to −0·9	2·0
−0·9 −0·8	5·0
−0·8 −0·7	7·0
−0·7 −0·6	15·0
−0·6 −0·5	19·0
−0·5 −0·4	12·0
−0·4 −0·3	24·0
−0·3 −0·2	28·0
−0·2 −0·1	18·0
−0·1 −0·0	17·5
+0·0 +0·1	13·5
+0·1 +0·2	23·0
+0·2 +0·3	22·0
+0·3 +0·4	7·0
+0·4 +0·5	20·0
+0·5 +0·6	25·0
+0·6 +0·7	17·0
+0·7 +0·8	13·0
+0·8 +0·9	14·0
+0·9 +1·0	4·0
Total	306·0
Mean	0·0502
s^2	0·2341
s	0·4838
β_1	0·1053
β_2	1·9603

Table 5. *Frequency distribution of multiple correlations between series with $n = 30$*

$R_{i.jk}$	Frequency
+0·9 to +1·0	22·0
+0·8 +0·9	53·0
+0·7 +0·8	70·0
+0·6 +0·7	42·5
+0·5 +0·6	39·0
+0·4 +0·5	29·0
+0·3 +0·4	18·5
+0·2 +0·3	18·5
+0·1 +0·2	8·5
+0·0 +0·1	5·0
Total	306·0
Mean	0·6314
s^2	0·4470
s	0·6686
β_1	0·4474
β_2	2·8261

In the second place, when we first discovered that the variance of the distribution of the zero order correlation does not become substantially smaller as n is increased* from 30 to 60 and 90, we thought this implied that little was to be gained by use of greater lengths of time series in forming estimates of inter-relationship. This, however, does not necessarily follow, even if it be granted that economic time series have approximately the autoregressive properties of our constructed series. In particular, it does not follow if we imagine that we are dealing with a relation between such time series in which the error term in the relation is a random variable of constant expected variance over time. In this case, the variance of the related series will continue to grow with time since there is no true mean, but the variance of the error term will not. Therefore, as the series become longer, the variance of the error term will become a smaller and smaller fraction of the variance of the series being explained. This implies that the correlation coefficient will become higher and higher approaching unity as the series approach an infinite length. Thus, whilst almost as high correlations are to be

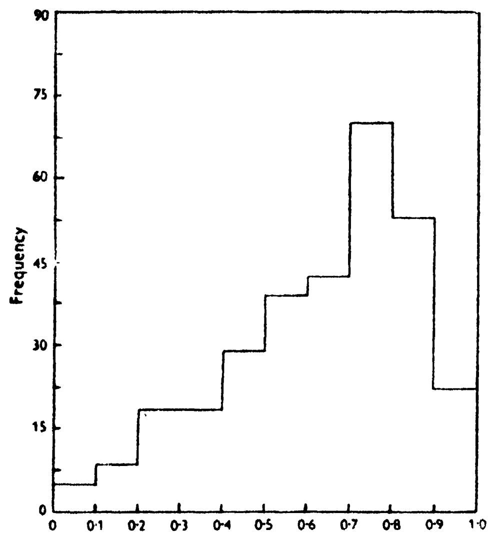


Fig. 6. Frequency distribution of $R_{i.jk}$, $n = 30$.

expected by chance between series of $n = 90$ as for $n = 30$, substantially higher correlations are to be expected as n increases if there is a real linear relation subject to a random error of constant expected variance over time. On the other hand, if the error term is not random in time but itself has continuity properties such as those of our series (6), then its expected variance will grow as the series become longer in the same way as the expected variance of our actual series grows. There will thus be no reason for the correlation to increase as the series become longer and we shall, in fact, be in the position of having gained little from the extension of the series. In this case consideration needs to be given to the possibility of correlating something like first differences in order to obtain any substantial advantage from the use of longer series.

In the third place, since for $n = 60$ and $n = 90$ we obtain good fits to our empirical distributions by means of the theoretical distributions of correlations between random series for

* We have received a very interesting letter from R. C. Geary in which he shows that for independent series generated by the rather similar process, $Y_t = Y_{t-1} + \epsilon_t$, the variance of correlations between series tends toward a non-zero finite positive quantity which cannot be very small.

$n = 5$ and $n = 6$, it follows that we might with some justification use these distributions for estimating the significance of correlations between economic time series. See David (1938) for tables of the correlation coefficient. Even in the case of $n = 30$, the use of the distribution for random series of $n = 5$ would not appear to overestimate the significance of correlations greater than 0·9.

Finally, it may be of some interest to note the result if we make use of the mean first lag autocorrelation of our 108 series with $n = 30$ and attempt to estimate var r by means of equation (1). However, instead of evaluating equation (1) as given, we reduced it to the following expression

$$\text{var } r \sim \frac{(1+r_1 r_1')}{n(1-r_1 r_1')} - \frac{2r_1 r_1'(1-r_1^n r_1'^n)}{n^2(1-r_1 r_1')^2}. \tag{7}$$

We have also substituted r_1 for ρ_1 and r_1' for ρ_1' since we intend to use sample rather than theoretical values of the autocorrelations. Then using $1-\frac{1}{2}$ (the mean δ^2/s^2 as given on Table 1) for r_1 and r_2 we obtain var $r = 0{\cdot}2759$. This estimate should be compared with the variance 0·2736 of our empirical distribution given in Table 2, and considering that Bartlett assumed very long stationary series generated by a simple Markoff scheme, it is remarkable that the estimate of var r should be as close as it is.

III. Investigation of the relation of var r to the sample values of the first lag autocorrelations

For the purpose of this study the 108 series, with $n = 30$, were separated into seven groups on the basis of the observed values of their first lag autocorrelation, r_1. The definition of the autocorrelation coefficient used for this purpose was

$$r_1 = 1 - \tfrac{1}{2}\delta^2/s^2, \tag{8}$$

where $$\delta^2 = \sum_1^{n-1} (Y_{i+1} - Y_i)^2/(n-1), \quad \text{and} \quad s^2 = \sum_1^{n} (Y_i - \bar{Y})^2/n.$$

As already mentioned, the distribution of the values of δ^2/s^2 for our 108 series is given in Table 1. The limits of r_1, the mean value of r_1, and the number of series for each of the seven groups of series is given in Table 6. The limits were chosen so as to obtain approximately equal numbers of series in each group.

Table 6. *Classification of series on the basis of sample values of first lag autocorrelations*

Group no.	Limits of the r_1 included in group	$\bar{r}_1$ for group	No. of series in group
1	0·570–0·832	0·732	15
2	0·832–0·887	0·866	15
3	0·887–0·912	0·896	15
4	0·912–0·946	0·927	18
5	0·946–0·963	0·955	18
6	0·963–0·975	0·970	14
7	0·975–0·988	0·980	13

The set of 1890 correlations, for $n = 30$, were sorted into 28 classes corresponding to the combinations of groups from which the paired series yielding the correlations were drawn. The variance of these correlations about zero is given for each of the 28 classes by the top

number in each box of Table 7. The middle figure in each of these boxes is the number of correlations that occurred in the class. The bottom figure in each class is a 'theoretical variance' obtained by using equation (7) with r_1 and r_1' taken as the mean values of the first lag sample autocorrelations corresponding to the groups of series being paired. For example, to estimate the variance of correlations between series from set 1 and from set 3, we used

Table 7. *Actual and theoretical values of* var r *for each of the* 28 *classifications of* r *based on the sample autocorrelations of the paired series**

		Group no. of first series of each pair						
		1	2	3	4	5	6	7
Group no. of second series of each pair	1	0·054 34 0·104	0·121 75 0·138	0·101 72 0·148	0·110 78 0·159	0·154 98 0·170	0·146 64 0·178	0·160 23 0·183
	2		0·188 31 0·206	0·159 72 0·232	0·213 86 0·259	0·213 92 0·300	0·263 73 0·312	0·267 69 0·324
	3			0·135 37 0·259	0·209 102 0·295	0·274 81 0·338	0·243 69 0·366	0·340 56 0·375
	4				0·308 57 0·343	0·282 89 0·403	0·316 98 0·440	0·352 61 0·457
	5					0·403 52 0·464	0·459 76 0·532	0·601 90 0·632
	6						0·510 31 0·616	0·599 49 0·637
	7							0·784 27 0·701

* The top number in each box is the actual variance, the middle number is the number of correlations in the class, and the lower number in each box is a 'theoretical' variance.

0·732 for r_1 in equation (7) and 0·896 for r_1'. It will be noticed that equation (7) only differs from the more approximate form of equation (2) in the right-hand term. For low values of r_1 or r_1' or large n, this term will be insignificant. Thus, whereas var r, as estimated by equation (7) for class 1–1, is only 0·006 less than the same estimate made by equation (2), we find that for class 4–4 the use of equation (7) gives an estimate which is 0·097 less than that obtained using equation (2). For higher values of $r_1 r_1'$ the difference becomes rapidly greater.

In Fig. 7 we have plotted the observed value of the variance for each class against the 'theoretical' value as obtained using equation (7). The straight line on this diagram is merely the 45° slope and would hold if the 'theoretical' variance agreed exactly with the actual. Notwithstanding some evidence of a systematic disagreement, the agreement obtained is remarkably good when it is considered that the assumptions under which Bartlett obtained the formula used for var r are far from being realized in this case. Not only are we dealing with non-stationary series having a small n, but we have used sample autocorrelations rather than true autocorrelations, and in addition our series were not generated by a Markoff process. Evidently we shall not go far wrong, even with our type of series, if for purposes of testing the null hypothesis we estimate the variance of a correlation between two series by using equation (7). The error that this involves for series of our type would appear, on the basis of Fig. 7, to be an overestimation of the variance to the extent of about 15 % for almost the entire range of sample first lag autocorrelations covered by our experiment.

Table 8. *Frequency distributions of the absolute values of correlations between series grouped according to their sample autocorrelations*

$\|r\|$	A	B	C	D	E	F	G
	Classes 1–1, 1–2, 1–3, 1–4	Classes 1–5, 1–6, 1–7, 2–2	Classes 2–3, 2–4, 3–3, 3–4	Classes 2–5, 2–6, 2–7, 3–5	Classes 4–4, 3–6, 3–7, 4–5	Classes 4–6, 4–7, 5–5, 5–6	Classes 6–6, 5–7, 6–7, 7–7
0 –0·10	56·50	56·25	37·00	29·75	14·50	19·75	1·00
0·10–0·20	60·25	37·00	45·00	33·50	20·50	21·75	2·25
0·20–0·30	45·75	43·50	40·50	29·25	32·00	19·50	4·00
0·30–0·40	30·75	35·50	41·50	34·00	30·75	17·75	3·00
0·40–0·50	28·00	27·25	41·50	49·50	28·75	24·00	5·25
0·50–0·60	21·50	25·50	37·25	48·50	47·50	32·00	12·00
0·60–0·70	12·25	18·00	30·25	38·25	44·75	37·75	24·75
0·70–0·80	4·00	19·25	15·00	36·50	32·75	61·00	44·25
0·80–0·90	0·00	3·75	9·00	15·75	17·75	48·00	58·25
0·90–0·95	0·00	0·00	0·00	0·00	1·75	5·50	31·25
0·95–1·00	0·00	0·00	0·00	0·00	0·00	0·00	11·00
Sum	259	266	297	315	271	287	197
Mean-square about 0	0·1070	0·1595	0·1889	0·2527	0·2885	0·3776	0·6138
β_2 (using moments about zero)	2·447	2·283	1·619	1·674	1·559	1·462	1·140

Before the above method for estimating var r can be fully utilized, it is necessary to know something about the shape of the distributions involved and, in order to provide information about this point, we obtained the frequency distributions tabulated in Table 8 and shown in Figs. 8–14. The 28 classes of correlations were ordered according to the theoretical values of var r as given in Table 7. Then the correlations of the first four classes were grouped together into class A, the second four classes were grouped together into class B, and so on. Because these distributions should be symmetrical about zero and in view of the small sample sizes, we have obtained frequency distributions of absolute values of r. These distributions make it clear that the distribution of r for fixed values of ρ_1 and ρ_1' depends on r_1 and r_1' to a very

considerable extent. It is rather interesting to note the way in which the mode gradually moves from zero towards 1 as the variance of the distribution increases. It is also interesting to note that, even for very high variances such as for distributions F and G, the ordinates of the curves still appear to approach zero for r approaching unity.

We were interested in seeing how well our empirical frequency distributions could be fitted by frequency distributions of approximately the same variance of correlations between series of normally distributed random items. Since the variance of correlations between random series is $1/(n-1)$, we have used the distribution for random series of $n = 11$ to fit the frequency distribution of A, Fig. 10, the distribution for random series of $n = 7$ for B, that of $n = 6$ for C, and that of $n = 5$ for D. We did not bother with the rest since it is obvious that the fit would be completely unsatisfactory. As a rough test of the goodness of the above fits, we applied a χ^2 test in each case. In case A we obtained a χ^2 of 2·663 with 6 d.f. and this corresponds to a probability of above 0·8. In case B, χ^2 was 5·105 with 6 d.f. and this corresponds to a probability of about 0·5. For C, χ^2 was 9·082 with 8 d.f. and this corresponds to a probability of about 0·3. In case D, χ^2 was 33·634 with 8 d.f. and this has a probability of less than 0·001. The same remarks apply concerning the roughness of the above test as we made in § II and it should also be true here that our use of the χ^2 test tends to underestimate the probabilities.

The primary implication of the results given in this section appears to us as follows. A reasonable way of testing the significance on the null hypothesis of a correlation between economic time series is first to estimate var r by means of equation (7) and, in so doing, to use the sample values of the first lag autocorrelations of the two series. Secondly, if the estimated var r is less than about 0·25, then insert it into the formula var $r = 1/(n'-1)$ and evaluate n'. Round n' off to the nearest integer and make use of the distribution of r that applies for random series with n equal to this integer to evaluate the probability that r might have occurred by chance between two non-related series. If the estimated var r turns out to be more than 0·25, then it seems unlikely that this method is valid, but the above test of significance would appear to be stronger than necessary and so would at least be a safe one to apply.

Secondly, our study would seem to offer some evidence that the variance of correlations between pairs of unrelated series of a given n, and with given sample autocorrelations, will be nearly if not completely independent of the true autocorrelations of the series. This follows from the fact that in Fig. 9 the points follow very closely the line that they would have been expected to follow if we had been dealing with sets of correlations associated with sets of series having different true autocorrelations. If the true autocorrelations were actually known, one might, at least in principle, first evaluate the probability of independently drawing two series with certain sample autocorrelations, and secondly evaluate the probability of obtaining by chance a given correlation between series having their sample autocorrelations. Having done this, it might be possible to evaluate the combined probability of drawing two series independently which had the autocorrelations and correlation between them which were actually obtained.

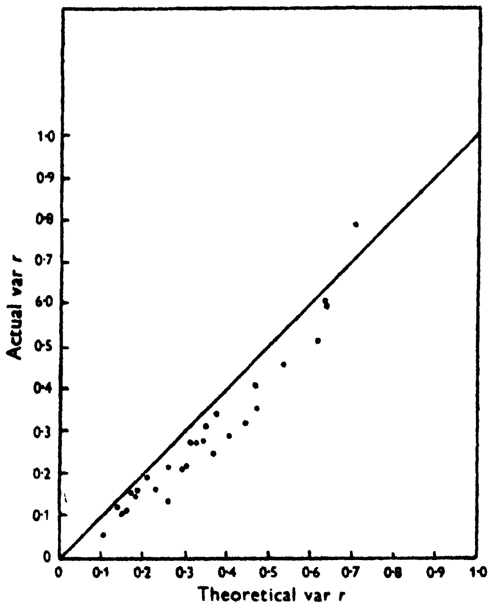


Fig. 7.

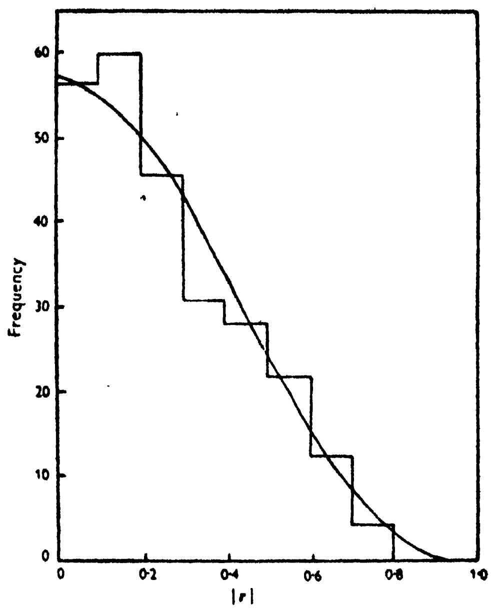


Fig. 8A.

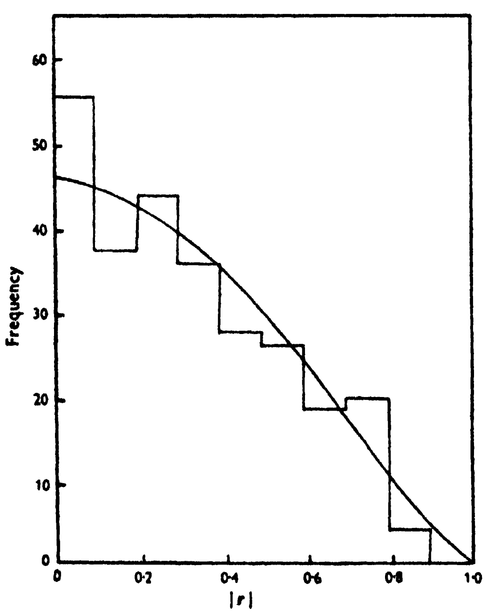


Fig. 9B.

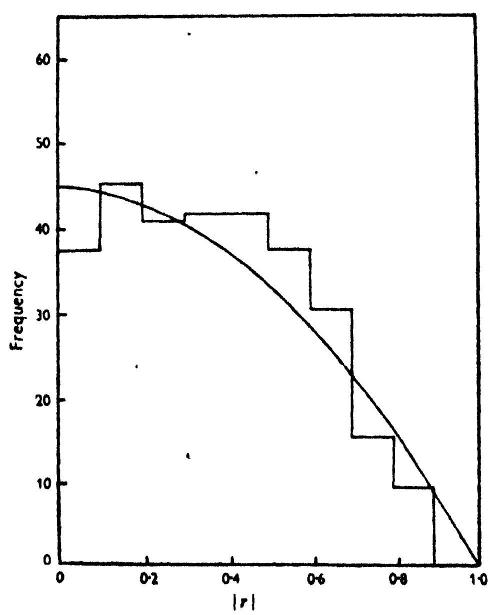


Fig. 10C.

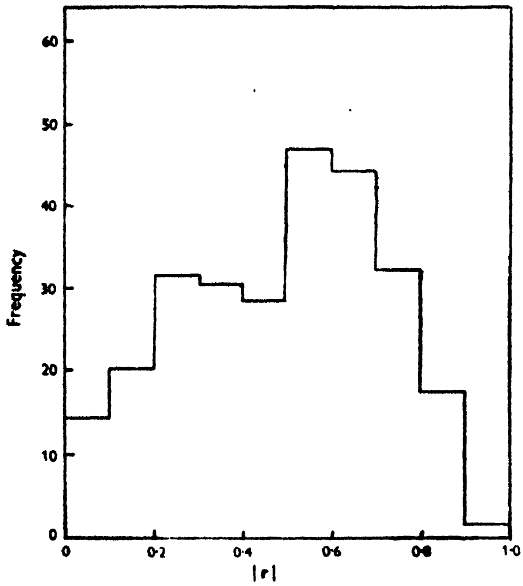


Fig. 11 D.

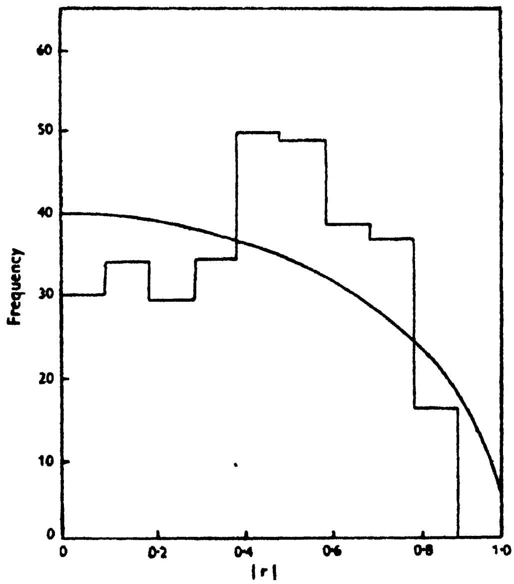


Fig. 12 E.

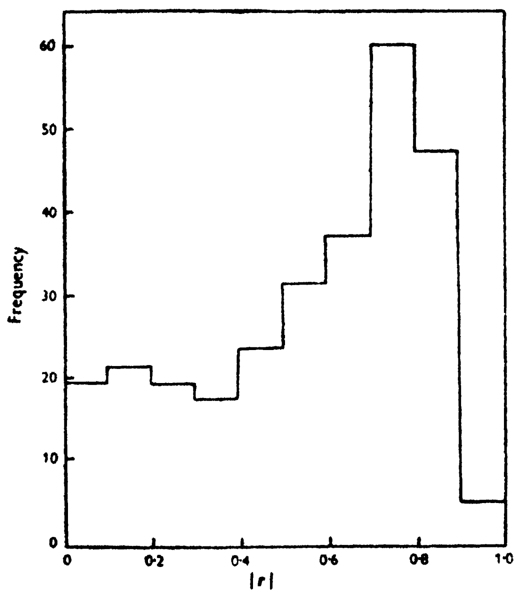


Fig. 13 F.

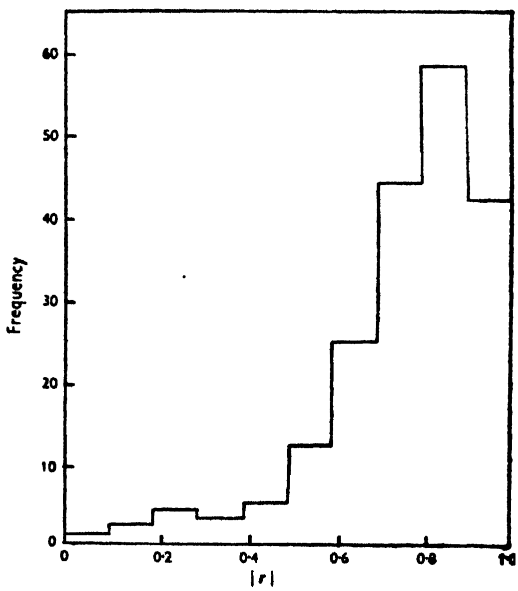


Fig. 14 G.

IV. Summary and a general remark on the problem of detecting real relations

The problem dealt with in this paper was that of testing the significance of correlations between economic time series. We constructed a set of non-related series with the same properties as are believed to hold for yearly series of a substantial group of economic time series. We then obtained a large number of correlations between these constructed series and on the basis of various distributions of these correlations came to the following conclusions.

1. On the assumption that the economic time series being dealt with are analogous to our specific model, $Y_{t+1} = Y_t + 0 \cdot 3(Y_t - Y_{t-1}) + \epsilon_t$, correlations between pairs of series can be tested for the null hypothesis using our empirical distributions. Since it was shown that for $n = 60$ and $n = 90$ good fits to our empirical distributions can be obtained by use of the ordinary distribution of correlations between random series with $n = 5$ and $n = 6$ respectively, it follows that an alternative procedure is to test correlations between economic series by means of these distributions.

2. The distribution of correlations between non-related series depends primarily on the sample autocorrelations of the paired series and very little, if at all, on the true autocorrelations, given the sample values. This makes it reasonable to apply tests based upon sample autocorrelations, and it was shown that one reasonable procedure is to use equation (7) to estimate the variance of r and then use the estimated variance of r to select the appropriate distribution of correlations between random series. Having chosen this distribution, one can then test the significance of the correlation in the usual way. The properties of this conditional test might repay a theoretical examination.

3. If economic time series are analogous to the constructed series used in this paper then, except in the cases where the sample autocorrelations happen to be low, such high correlations between economic time series may be expected by chance that we are unlikely to detect real relations. The distributions given for the partial and multiple correlations only accentuate this view. One method which suggests itself of at least partially extricating ourselves from this situation is to make an autoregressive transformation of the time series involved in such a way that at least one of the series becomes approximately random in time. When this has been done, the correlation between the transformed series may then be tested in the usual way (Bartlett, 1935). Thus, for example, suppose that

$$x_{1t} = \beta x_{2t} + u_t, \tag{9}$$

and the error term u_t is generated by

$$u_t = \alpha u_{t-1} + \epsilon_t, \tag{10}$$

where ϵ_t is a random variable. If $\beta = 0$, then $x_{1t} = u_t$ and an appropriate autoregressive transformation is

$$x'_{1t} = x_{1t} - \alpha x_{1(t-1)}, \quad x'_{2t} = x_{2t} - \alpha x_{2(t-1)},$$

so that in terms of the transformed variables we have

$$x'_{1t} = \beta x'_{2t} + \epsilon_t. \tag{11}$$

Since one of the two variables is now random we can apply the usual test of significance of correlation between x'_{1t} and x'_{2t} with the consequent advantage of a great increase in the effective degrees of freedom. On the other hand, if $\beta \neq 0$, then the expected value of the least square estimate of β in equation (11) will still be the same as in equation (9) and the expected

value of estimates of the true correlation between x'_{1t} and x'_{2t} will still be very nearly the same as the expected value of estimates of correlation between x_{1t} and x_{2t}. This means that in this case, at least, the transformed form (11) is far more effective than the untransformed form (9), for the purpose of testing the null hypothesis. When $\beta \neq 0$, then the appropriate autoregressive transformation for estimating β is not one which leaves one of the series random but rather one that leaves the error term random. However, in the case $\beta = 0$, it is evident that when the error term is random then the variable which is taken to be dependent must also be random since it is entirely composed of the error term.*

REFERENCES

AITKEN, A. C. (1935). On least squares and linear combination of observations. *Proc. Roy. Soc. Edinb.* **55**, 42.

BARTLETT, M. S. (1935). Some aspects of the time correlation problem in regard to tests of significance. *J. Roy. Statist. Soc.* **98**, 536.

BARTLETT, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time-series. *J. Roy. Statist. Soc.* Supplement, **8**, 27.

BARTLETT, M. S. (1947). *Stochastic Processes.* Mimeographed notes of a course given at the University of North Carolina in the Fall Quarter, 1946.

CHAMPERNOWNE, D. C. (1948). Sampling theory applied to autoregressive time-series. *J. Roy. Statist. Soc., Series B*, **10** (in press).

COCHRANE, D. & ORCUTT, G. H. (1949). Application of least squares regression analysis to relationships containing autocorrelated error terms. *J. Amer. Statist. Ass.* **44** (in press).

DAVID, F. N. (1938). *Tables of the Correlation Coefficient.* Cambridge University Press.

HART, B. I. & VON NEUMANN, J. (1942). Tabulation of the probabilities for the ratio of the mean square successive difference to the variance. *Ann. Math. Statist.* **13**, 207.

KENDALL, M. G. & BABINGTON SMITH, B. (1939). *Tracts for Computers*, no. XXIV. Cambridge University Press.

KENDALL, M. G. (1944). Oscillatory movements in English agriculture. *J. Roy. Statist. Soc.* **106**, 91.

KENDALL, M. G. (1945). On the analysis of oscillatory time-series. *J. Roy. Statist. Soc.* **108**, 93.

KENDALL, M. G. (1946). *Contributions to the Study of Oscillatory Time-Series.* National Institute of Economic and Social Research, Occasional Papers X. Cambridge University Press.

MORAN, P. A. P. (1947). Some theorems on time series. I. *Biometrika*, **34**, 281.

ORCUTT, G. H. (1948*a*). A study of the autoregressive nature of the time series used for Tinbergen's model of the economic system of the United States, 1919–1932. *J. Roy. Statist. Soc., Series B*, **10**, 1.

ORCUTT, G. H. (1948*b*). A new regression analyser. *J. Roy. Statist. Soc., Series A*, **111** (in press).

PEARSON, E. S. (1931). The test of significance for the correlation coefficient. *J. Amer. Statist. Ass.* **26**, 128.

QUENOUILLE, M. H. (1947). Notes on the calculation of autocorrelations of linear autoregressive schemes. *Biometrika*, **34**, 365.

TINBERGEN, J. (1939). *Statistical Testing of Business-Cycle Theories. Vol. 2: Business Cycles in the United States of America* 1919–1932. Geneva; League of Nations.

VON NEUMANN, J. (1941). Distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Statist.* **12**, 367.

VON NEUMANN, J. (1942). A further remark on the distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Statist.* **13**, 86.

WOLD, H. (1938). *A Study in the Analysis of Stationary Time Series.* Uppsala.

YULE, G. U. (1921). On the time correlation problem. *J. Roy. Statist. Soc.* **84**, 497.

YULE, G. U. (1926). Why do we sometimes get nonsense correlations between time series? *J. Roy. Statist. Soc.* **89**, 1.

YULE, G. U. (1927). On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers. *Philos. Trans.* A, **226**, 267.

* For a fuller discussion of the problem of estimation when the error term is autocorrelated, see Aitken (1935), Champernowne (1948) and Cochrane & Orcutt (1949).

MISCELLANEA

Note on the median of a multivariate distribution

By J. B. S. HALDANE

The median of a univariate distribution is an exceedingly useful parameter but, whereas the notions of the mean and mode can be applied without ambiguity to distributions in two or more dimensions, this is not so for the median. It is the object of this note to point out that when we are dealing with multivariate distributions, there are two quite distinct sets of parameters, each of which possess some of the properties of the univariate median, while lacking others.

The possibility arises from the fact that the univariate median is a location parameter associated with two quite different scale parameters. In the first place, for the distribution $dF = f(x)\,dx$, the median is defined as M, where

$$\int_{-\infty}^{M} dF = \int_{M}^{\infty} dF = \tfrac{1}{2}.$$

Integration here and throughout is understood in Stieltjes's sense.

When so defined, the median is obviously associated with the quartiles defined by $\int_{-\infty}^{Q_1} dF = \frac{1}{4}$ and $\int_{Q_3}^{\infty} dF = \frac{1}{4}$, and with the interquartile range $Q_3 - Q_1$.

Secondly, however, we may define the median as the value M which minimizes $\int_{-\infty}^{\infty} |M-x|\,dF$.

Similarly, the mean can be defined as the value m which minimizes $\int_{-\infty}^{\infty} (m-x)^2\,dF$. Now the minimum value of this quantity is simply the variance. Just as the mean is associated with the standard deviation as a measure of dispersion, so on this definition the median is associated with the mean deviation about the median. The more commonly used measure, the mean deviation about the mean, has perhaps less to recommend it, since it is not a stationary value, and therefore more liable to error if the corresponding scale parameter is in error. In geometrical language the median is the point the sum of whose distances from the representative points of the sample is a minimum.

Both these definitions of the median are equivalent in the univariate case, and both are of course indeterminate if the number of members of a sample is even, unless an even number of them coincide with the median. The various devices which avoid this indeterminacy represent the median as a limit.

When we pass to two or more dimensions these two definitions are no longer equivalent, and it seems worth while to distinguish the two analogues of the univariate median as the arithmetic and geometric medians.

If we have a number of variates $x, y, z, \ldots$ the arithmetic median is the set of values $(X, Y, Z, \ldots)$, where $X, Y, Z, \ldots$ are the medians of $x, y, z, \ldots$ defined in either of the two above ways. When x, y, z, etc., are different in kind it is obviously the only reasonable generalization. It has the merit of being invariant, like the median, when any of the variates is replaced by a monotonic function of it. But it is not invariant under a rotation of axes.

For consider the arithmetical median of three coplanar points. If we take rectangular axes their co-ordinates are (x_1, y_1), (x_2, y_2), (x_3, y_3). Those of the median are (x_m, y_m), where x_m is the middle value of x_1, x_2, x_3 if they are all different, and the value of the two equal ones if two are equal. Hence as we rotate the axes we find that the position of the arithmetic median changes. Unless it coincides with one of the apices of the triangle, one of the sides must subtend a right angle at it. In fact, the locus consists of those arcs of the circles which have the sides of the triangle as diameters which lie within the triangle (see Fig. 1). If the triangle has a right or obtuse angle, this angle lies on the locus, as does the foot of the perpendicular from it on the opposite side. If the triangle is acute angled, it passes through the feet of the three perpendiculars. Similarly, the locus of the arithmetic median of the vertices of a tetrahedron consists of portions of spheres. For an odd number of more than three points in a plane, the arithmetic median may always be one of them, or its locus may consist of a series of circular arcs. For an even number it is of course indeterminate, unless one or other of the special conventions devised for the univariate case is used.

The geometric median is defined as the point such that the sum of its distances from the sample points is a minimum. It is invariant under a change of axes, but is not invariant when the scales in different directions are altered. Its sole value is therefore in problems of geometrical probability. It occurred in a problem of this type during the recent war, and might perhaps be of value in studies on such aggregates as star clusters, where we desire to find a representative point which is less affected than the centroid by outliers which may not be members of the cluster.

The geometrical median of three coplanar points is the point in the triangle formed by them at which each side subtends an angle of $\frac{2}{3}\pi$, provided that no angle of the triangle exceeds $\frac{2}{3}\pi$. If one angle exceeds this value, the geometrical median is the obtuse vertex of the triangle. I have been unable to find any simple geometrical construction in the case of more than three points. It is, however, easy to show that

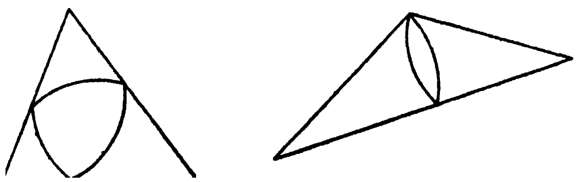


Fig. 1.

the geometrical median is unambiguously defined. For let us take it (or *per impossibile*, one of the geometrical medians) as our origin of Cartesian co-ordinates. Consider a set of n coplanar points (x_r, y_r). First suppose that no sample point coincides with the origin, and if necessary rotate the axes so that no axis passes through a sample point. Let R be the sum of the distances of the sample points from the point $(x, 0)$. Then

$$R = \sum_{r=1}^{n} [(x-x_r)^2+y_r^2]^{\frac{1}{2}}, \quad \frac{dR}{dx} = \sum_{r=1}^{n} [(x-x_r)\{(x-x_r)^2+y_r^2\}^{-\frac{1}{2}}].$$

This must be zero when $x = 0$. But

$$\frac{d^2R}{dx^2} = \sum_{r=1}^{n} [y_r^2\{(x-x_r)^2+y_r^2\}^{-\frac{3}{2}}].$$

All the terms in this sum are necessarily positive, since the denominator is the cube of a distance which is taken as positive and can never change its sign. Hence d^2R/dx^2 is always positive, and R has only one minimum.

Next suppose that the median coincides with one of the points, say the first; then

$$R = |x| + \sum_{r=2}^{n} [(x-x_r)^2+y_r^2]^{\frac{1}{2}}, \quad \frac{dR}{dx} = \pm 1 + \sum_{r=2}^{n} [(x-x_r)\{(x-x_r)^2+y_r^2\}^{-\frac{1}{2}}],$$

d^2R/dx^2 is positive as before, but dR/dx has a saltus at $x = 0$, increasing in value by 2, and changing sign. R has therefore a sharp minimum. The proof in three or more dimensions is analogous. Changing to polar co-ordinates with the origin as centre and the co-ordinates of the sample points as (ρ_r, θ_r), it follows that if the median does not coincide with one of them,

$$\Sigma \cos\theta_r = \Sigma \sin\theta_r = 0,$$

whilst if it does so, these sums lie between ± 1. If several sample points coincide with the geometric median, the modifications are obvious.

It is clear that the minimum sum of the distances, divided by the number of points, is the many-dimensional analogue of the mean deviation from the median in one dimension.

To sum up, the arithmetical median is obviously to be preferred in ordinary statistical work, but the geometrical median has certain advantages in problems of geometrical probability. In either case it is desirable to state clearly how the median is defined.

A property of rank correlations

By H. E. DANIELS

The product-moment correlation coefficient may be considered a satisfactory measure of the association between two variates, both because of its special relevance to the bivariate normal distribution, and, more generally, from the fact that the square of its population value is the fraction of the variance of one variate removed by linear regression on the other.

When the data are presented in ranked form, however, the suitable choice of a measure of concordance between rankings is less obvious. Spearman's analogous use of the product-moment correlation coefficient ρ between the ranks cannot be so easily justified, though when calculated from the sum of squares of rank differences it is seen to measure in a rather arbitrary sense the degree of agreement between ranks. Kendall's coefficient τ has a more direct interpretation in that it is a function of the total number of corresponding pairs of ranks which agree in order, and Moran (1948) has further shown that τ is simply related to the least number of interchanges required to bring two rankings into perfect agreement. It may be said, therefore, that τ is a satisfactory measure of concordance in the sense that increasing values of τ correspond to increasing agreement between the rankings as defined in either of these ways.

In a previous paper (1944) I introduced the quantity

$$\Gamma = \frac{\Sigma a_{ij} b_{ij}}{\sqrt{(\Sigma a_{ij}^2 \Sigma b_{ij}^2)}}$$

as a general measure of the correlation between two sets of observations, ranked or otherwise, where a_{ij}, b_{ij} are scores assigned to corresponding pairs i, j of the two variables and $a_{ij} = -a_{ji}$, $b_{ij} = -b_{ji}$. Both ρ and τ are included as special cases. It is of interest to see how far Γ may be justified as a suitable measure of rank correlation.

Provided the scores have the same sign as the difference of the corresponding ranks, Γ takes the values ± 1 when there is respectively complete concordance or discordance between the rankings, and when the ranked attributes are independent, Γ is on the average zero (though it must be remembered that a zero average value of Γ does not necessarily imply independence). A further property is, however, required before Γ can be accepted as satisfactory; it has to be shown that increasing values of Γ correspond in some way to increasing concordance between the rankings.

A property of this type which one might expect of a rank correlation coefficient is that if any two corresponding pairs of ranks do not agree in order, the value of the coefficient should increase when the members of one of the pairs are interchanged. It is now shown that Γ has this property provided the scores a_{ij}, b_{ij} do not decrease with increasing rank separation and are not zero except for tied ranks. The scores for both ρ and τ satisfy the conditions.

Let p_i, q_i be the ranks of the ith members of the two sets of observations, and suppose that $p_r > p_s$, $q_r < q_s$, and that p_r, p_s are to be interchanged. The denominator of Γ is unaffected by any relative permutation of the ranks. Initially, the numerator is

$$\begin{aligned}\Sigma a_{ij} b_{ij} &= \Sigma'' a_{ij} b_{ij} + \Sigma'' a_{rj} b_{rj} + \Sigma'' a_{ir} b_{ir} + \Sigma'' a_{sj} b_{sj} + \Sigma'' a_{is} b_{is} + a_{rs} b_{rs} + a_{sr} b_{sr} \\ &= \Sigma'' a_{ij} b_{ij} + 2\Sigma'' a_{rj} b_{rj} + 2\Sigma'' a_{sj} b_{sj} + 2a_{rs} b_{rs},\end{aligned}$$

where Σ'' denotes summation over all values of i, j excluding r and s. After interchanging p_r, p_s it becomes

$$\Sigma'' a_{ij} b_{ij} + 2\Sigma'' a_{sj} b_{rj} + 2\Sigma'' a_{rj} b_{sj} - 2a_{rs} b_{rs},$$

which is an increase of

$$-2\Sigma''(a_{sj} - a_{rj})(b_{sj} - b_{rj}) - 4a_{rs} b_{rs} = -2\Sigma(a_{sj} - a_{rj})(b_{sj} - b_{rj}).$$

Now introduce the condition that a_{ij} is a non-decreasing function of $p_j - p_i$, and consider $a_{sj} - a_{rj}$ for all values of j. The following are the possible alternatives.

(i) $p_j \geq p_r > p_s$. $a_{sj} > 0$, $a_{rj} \geq 0$ and $a_{sj} \geq a_{rj}$.
(ii) $p_r > p_j \geq p_s$. $a_{sj} \geq 0$, $a_{rj} < 0$, so that $a_{sj} - a_{rj} = a_{sj} + a_{jr} > 0$.
(iii) $p_r > p_s > p_j$. $a_{sj} < 0$, $a_{rj} < 0$ and $a_{jr} \geq a_{js}$, so that $a_{sj} \geq a_{rj}$.

Thus in all cases $a_{sj} - a_{rj} \geq 0$ and in at least one case $a_{sj} - a_{rj} > 0$.

Similarly, since $q_r < q_s$ the fact that b_{ij} is a non-decreasing function of $q_j - q_i$ implies that $b_{sj} - b_{rj} \leq 0$ and < 0 at least when $a_{sj} - a_{rj} > 0$. It follows that Γ is increased by the interchange.

On the other hand, if the scores decrease with increasing rank separation the result is not necessarily true. For suppose $a_{ij} = \pm A$, a large number, when $p_j - p_i = \pm 1$, and $a_{ij} = \pm 1$ as for Kendall's τ when $|p_j - p_i| > 1$, with similar scores b_{ij} for $q_j - q_i$. Then for the ranking

$$\begin{array}{cccccccc} 8 & 6 & 7 & 1 & 4 & 2 & 5 & 3 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array}$$

the value of Γ is decreased on interchanging 6 and 4 in the first ranking.

When tied ranks are present, any pair of ties is scored zero, but some consideration will show that Γ is still increased if a discordant pair is brought to concordance by an interchange.

The interchange of a particular pair of discordant ranks will in general alter the order of some of the other pairs which involve one or other of the ranks interchanged, but it is worth remarking that by virtue of the result just proved the value of τ, and hence the total number of concordances between pairs of ranks, must be increased.

REFERENCES

MORAN, P. A. P. (1948). *Proc. Camb. Phil. Soc.* **44**, 142.
DANIELS, H. E. (1944). *Biometrika*, **33**, 129.

Approximation errors in distributions of independent variates

BY H. O. HARTLEY

1. Let x and y be two independent variates and let $q = \phi(x, y)$ be a function of these variates (e.g. the ratio x/y or the product $x.y$). In distribution theory one is frequently faced with the following problem:

If we approximate to the distribution of x and/or y, what is the effect of such approximations on the distribution of q?

In this note we derive a lemma which, although mathematically trivial, provides a gauge for this error and has been of great help in recent work. It appeared worth while therefore to put it on record (see § 5).

2. To fix the ideas we assume that $0 < x < \infty$, $0 < y < \infty$, and that the differentiable function $q = \phi(x, y)$ is monotonic increasing in x and monotonic in y, so that differentiable and monotonic inversion functions $x = \psi(q, y)$ and $y = \chi(q, x)$ exist. It is obvious from the argument given below that some of these restrictions are in fact not essential.

3. Let $F(X)$ be the chance for $0 \leqslant x \leqslant X$ and $G(Y)$ be the chance for $0 \leqslant y \leqslant Y$, and denote by $f(X)$ and $g(Y)$ approximations to F and G. We shall assume that these approximations are themselves 'probability integrals', so that

$$\lim_{X\to\infty} f(X) = \lim_{Y\to\infty} g(Y) = 1 \quad \text{and} \quad f'(X) \geqslant 0, \quad g'(Y) \geqslant 0, \tag{1}$$

where the distribution functions f' and g' are the differentials of f and g. Let us denote the differences between exact probability integrals and the approximations by

$$\epsilon(X) = F(X) - f(X), \quad \eta(Y) = G(Y) - g(Y), \tag{2}$$

so that

$$\lim_{X\to\infty} \epsilon(X) = \lim_{Y\to\infty} \eta(Y) = 0, \quad \epsilon(0) = \eta(0) = 0. \tag{3}$$

The probability integral of q, i.e. the chance ($H(Q)$ say) of $q = \phi(x, y) \leqslant Q$, is then given by

$$H(Q) = \underset{\phi(x,y)\leqslant Q}{\int_0^\infty \int_0^\infty} F'(x)\, G'(y)\, dx\, dy, \tag{4}$$

where F' and G' are the distribution functions of x and y. We have from (2) and (4)

$$\begin{aligned} H(Q) &= \int_0^\infty \int_0^\infty f'(x)\, g'(y)\, dx\, dy + \int_0^\infty \int_0^\infty \{F'\eta' + g'\epsilon'\}\, dx\, dy \Bigg| \\ &\quad \phi(x,y) \leqslant Q \qquad\qquad\qquad\qquad \phi(x,y) \leqslant Q \\ &= h(Q) \qquad\qquad\qquad\qquad\quad + \theta(Q) \quad \text{(say)}. \end{aligned} \tag{5}$$

The integral $h(Q)$ is the approximation to $H(Q)$ computed from the approximate probability integrals f and g and $\theta(Q)$ the error thereby committed.

4. To estimate this error we have

$$\theta(Q) = \int_0^\infty \left\{\eta'(y) \int_0^{\psi(Q,y)} F'(x)\, dx\right\} dy + \int_0^\infty \left\{g'(y) \int_0^{\psi(Q,y)} \epsilon'(x)\, dx\right\} dy.$$

Applying partial integration to the first term, ordinary integration to the second term and noting (3), we reach

$$\theta(Q) = -\int_0^\infty \left\{F'(\psi(Q,y))\, \eta(y) \frac{\partial \psi}{\partial y}(Q,y)\right\} dy + \int_0^\infty \{g'(y)\, \epsilon(\psi(Q,y))\}\, dy$$

or

$$\theta(Q) = -\int_{\psi(Q,\ )}^{\psi(Q,\infty)} F'(\psi)\, \eta(\chi(Q,\psi))\, d\psi + \int_0^\infty \{g'(y)\, \epsilon(\psi(Q,y))\}\, dy.$$

Now since $F' \geqslant 0$ and $g' \geqslant 0$, and since

$$\left| \int_{\psi(Q,0)}^{\psi(Q,\infty)} F'(\psi)\, d\psi \right| \leqslant \left| \int_0^\infty F'(\psi)\, d\psi \right| = 1, \quad \int_0^\infty g'(y)\, dy = 1,$$

we obtain immediately $\qquad |\theta(Q)| \leqslant \max|\eta(y)| + \max|\epsilon(x)|,$

and this inequality proves the following lemma:

5. Lemma. Let x and y be independent variates with probability integrals $F(X)$ and $G(Y)$ respectively; let $f(X)$ and $g(Y)$ be approximations to F and G with errors $\epsilon(X)$ and $\eta(Y)$ respectively; finally, let $q = \phi(x, y)$ be a function of the variates x and y satisfying the above conditions (see §2). If, then, the probability integral of q is computed from f and g, the error thereby committed is smaller than $\max|\epsilon| + \max|\eta|$. By repeated application of the lemma, the generalization to functions of more than two variates, $q = \phi(x_1, x_2, \ldots x_m)$, is obvious.

Correlations between χ^2 cells

By F. N. DAVID

1. The population studied is assumed to fall into k groups or strata, there being a proportion p_i ($i = 1, 2, \ldots, k$) in the ith group. A sample of size N is randomly and independently drawn from the population, the number falling into the ith group being n_i. We write

$$m_i = Np_i \quad \text{and} \quad x_i = n_i - m_i.$$

It is well known that if the only restriction which is placed on the sample is that the totals of observation and expectation are made to agree, i.e. if

$$\sum_{i=1}^{k} x_i = 0,$$

then, writing r_{ij} for the coefficient of correlation between x_i and x_j,

$$r_{ij} = -\left(\frac{m_i m_j}{(N-m_i)(N-m_j)}\right)^{\frac{1}{2}}.$$

2. The coefficient of correlation between x_i and x_j when more than one restriction is placed on the sample is not easily determined. It may, however, be deduced from a consideration of the multivariate normal surface. The multivariate normal surface may be written

$$p(x_1 x_2 \ldots x_n) = \frac{1}{(2\pi)^{\frac{1}{2}n}\, \sigma_1 \sigma_2 \ldots \sigma_n R^{\frac{1}{2}}} \exp\left\{-\frac{1}{2R}\left[\sum_{i=1}^{n} R_{ii} \frac{x_i^2}{\sigma_i^2} + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} R_{ij} \frac{x_i x_j}{\sigma_i \sigma_j}\right]\right\},$$

where
$$R = \begin{vmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1n} \\ r_{21} & 1 & r_{23} & \cdots & r_{2n} \\ r_{31} & r_{32} & 1 & \cdots & r_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{n1} & r_{n2} & r_{n3} & \cdots & 1 \end{vmatrix}.$$

R_{ij} is the minor obtained by omitting the ith row and the jth column, and r_{ij} is the coefficient of correlation between x_i and x_j. Consider the exponent, and write it as the sum of n linear squares, viz.:

$$(\alpha_{11}x_1+\alpha_{12}x_2+\alpha_{13}x_3+\ldots+\alpha_{1n}x_n)^2+(\alpha_{22}x_2+\alpha_{23}x_3+\ldots+\alpha_{2n}x_n)^2+\ldots+\alpha_{nn}^2x_n^2.$$

Make the substitutions
$$\begin{aligned} z_1 &= \alpha_{11}x_1+\alpha_{12}x_2+\ldots+\alpha_{1n}x_n, \\ z_2 &= \qquad\quad \alpha_{22}x_2+\ldots+\alpha_{2n}x_n, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ z_n &= \qquad\qquad\qquad\quad \alpha_{nn}x_n, \end{aligned}$$

and solve for the x's. We have then a series of relations of the form
$$\begin{aligned} x_1 &= \beta_{11}z_1+\beta_{12}z_2+\ldots+\beta_{1n}z_n, \\ x_2 &= \qquad\quad \beta_{22}z_2+\ldots+\beta_{2n}z_n, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_n &= \qquad\qquad\qquad\quad \beta_{nn}z_n. \end{aligned}$$

If we now put $x_1 = x_2 = \ldots = x_n = 0$, we have n equations which may be regarded as n planes in an n-dimensioned space. If θ_{ij} is the angle between the ith and the jth plane, then
$$\theta_{ij} = \cos^{-1} r_{ij}.$$

3. I believe this last relation to be well known,* although I have not been able to find a reference to its proof anywhere. The proof is, however, quite straightforward, making use of the well-known relations between the minors of a determinant. It is suggested that this result may be used to find the correlation between x_i and x_j of § 1. It is assumed that the number, N, in the sample, and the number of groups are large enough for the assumptions regarding χ^2 to be satisfied; that is to say, it is assumed that x_i is normally distributed about m_i for $(i = 1, 2, \ldots, k)$. We consider a χ^2 of k groups where we have placed p linear restraints on the x's. The probability $P\{\chi^2 > \chi_0^2\}$, where χ_0^2 is some constant, is given by the multiple integral in a $(k-p)$-dimensioned space taken over the domain D defined by $\chi^2 > \chi_0^2$. We have then
$$P\{\chi^2 > \chi_0^2\} = \text{constant} \times \int\int_D \ldots \int \exp\left[-\tfrac{1}{2}\sum_{i=1}^{k}\frac{x_i^2}{m_i}\right] dx_1 dx_2 \ldots dx_{k-p}, \tag{1}$$
where $x_{k-p+1}, \ldots, x_k$ can be expressed in terms of $x_1, x_2, \ldots, x_{k-p}$ by solving the equations by means of which the linear restraints are expressed. The expression under the integral sign is equivalent to a multivariate normal distribution in $k-p$ dimensions; accordingly, by writing the exponent in the way described in § 2 and finding the angle between the appropriate planes $x_i = 0$ and $x_j = 0$, r_{ij} can be deduced.

4. As an illustration consider the case of a population which is divided into five groups. A sample of N is drawn, and, using the notation of § 1, we have that the correlation between x_1 and x_2 is, for the case of one restraint,
$$r_{12} = -\left(\frac{m_1 m_2}{(N-m_1)(N-m_2)}\right)^{\frac{1}{2}}.$$

We begin by showing that this may be deduced by the method outlined here. If the one restraint is that the sum of the x's is zero, then we have
$$x_1+x_2+x_3+x_4+x_5 = 0 \quad \text{or} \quad -x_5 = x_1+x_2+x_3+x_4.$$

The exponent under the χ^2 integral may therefore be written (dropping the multiplier $-\frac{1}{2}$ for convenience)
$$\frac{x_1^2}{m_1}+\frac{x_2^2}{m_2}+\frac{x_3^2}{m_3}+\frac{x_4^2}{m_4}+\frac{(x_1+x_2+x_3+x_4)^2}{m_5}.$$

---

* It is easily deduced from determinantal relations given by K. Pearson in lectures and in various papers dealing with multiple correlation. M. G. Kendall points out that it follows as a natural result of the discussion on p. 372 of his *Advanced Theory of Statistics*, vol. 1. Sheppard used the bivariate result as a means of estimating the correlation coefficient.

This may be rewritten as

$$\left[x_1\left(\frac{m_1+m_5}{m_1m_5}\right)^{\frac{1}{2}}+x_2\left(\frac{m_1}{m_5(m_1+m_5)}\right)^{\frac{1}{2}}+x_3\left(\frac{m_1}{m_5(m_1+m_5)}\right)^{\frac{1}{2}}+x_4\left(\frac{m_1}{m_5(m_1+m_5)}\right)^{\frac{1}{2}}\right]^2$$
$$+\left[x_2\left(\frac{m_1+m_2+m_5}{m_2(m_1+m_5)}\right)^{\frac{1}{2}}+x_3\left(\frac{m_2}{(m_1+m_2+m_5)(m_1+m_5)}\right)^{\frac{1}{2}}+x_4\left(\frac{m_2}{(m_1+m_2+m_5)(m_1+m_5)}\right)^{\frac{1}{2}}\right]^2$$
$$+\left[x_3\left(\frac{N-m_4}{m_3(m_1+m_2+m_5)}\right)^{\frac{1}{2}}+x_4\left(\frac{m_3}{(N-m_4)(m_1+m_2+m_5)}\right)^{\frac{1}{2}}\right]^2+\left[x_4\left(\frac{N}{m_4(N-m_4)}\right)^{\frac{1}{2}}\right]^2.$$

Substituting z_1 for the expression in the first bracket, z_2 for the second and so on, and solving for x_1, x_2, x_3 and x_4 in terms of the z's, we have

$$x_1=\left(\frac{m_1m_5}{m_1+m_5}\right)^{\frac{1}{2}}z_1-\frac{m_1m_2^{\frac{1}{2}}}{(m_1+m_5)^{\frac{1}{2}}(m_1+m_2+m_5)^{\frac{1}{2}}}z_2-\frac{m_1m_3^{\frac{1}{2}}}{(m_1+m_2+m_5)^{\frac{1}{2}}(N-m_4)^{\frac{1}{2}}}z_3-\frac{m_1m_4^{\frac{1}{2}}}{N^{\frac{1}{2}}(N-m_4)^{\frac{1}{2}}}z_4,$$
$$x_2=\frac{(m_1+m_5)^{\frac{1}{2}}m_2^{\frac{1}{2}}}{(m_1+m_2+m_5)^{\frac{1}{2}}}z_2-\frac{m_2m_3^{\frac{1}{2}}}{(m_1+m_2+m_5)^{\frac{1}{2}}(N-m_4)^{\frac{1}{2}}}z_3-\frac{m_2m_4^{\frac{1}{2}}}{N^{\frac{1}{2}}(N-m_4)^{\frac{1}{2}}}z_4,$$
$$x_3=\frac{(m_1+m_2+m_5)^{\frac{1}{2}}m_3^{\frac{1}{2}}}{(N-m_4)^{\frac{1}{2}}}z_3-\frac{m_3m_4^{\frac{1}{2}}}{(N-m_4)^{\frac{1}{2}}N^{\frac{1}{2}}}z_4,$$
$$x_4=\frac{(N-m_4)^{\frac{1}{2}}m_4^{\frac{1}{2}}}{N^{\frac{1}{2}}}z_4.$$

The angle between the planes $x_1=0$ and $x_2=0$ will be

$$\theta_{12}=\cos^{-1}-\frac{m_1m_2}{N}\Big/\left(\frac{m_1(N-m_1)}{N}\right)^{\frac{1}{2}}\left(\frac{m_2(N-m_2)}{N}\right)^{\frac{1}{2}}=\cos^{-1}-\left(\frac{m_1m_2}{(N-m_1)(N-m_2)}\right)^{\frac{1}{2}}.$$

For this case the result may also be reached by calculating

$$\theta_{34}=\cos^{-1}-\left(\frac{m_3m_4}{(N-m_3)(N-m_4)}\right)^{\frac{1}{2}}$$

and deducing θ_{12} by symmetry.

5. It will be noted that the restraint placed on χ^2 in the preceding section is equivalent to making the totals of observation and expectation agree, as was pointed out in § 1. We now proceed to place such further restraints on χ^2 as are made when moments of observation and expectation are put in agreement. Admittedly this narrows the field of investigation to a certain extent, but it may be argued that such restraints are those which are most often met in practice. We assume therefore that the two restraints placed on the sample are

$$x_1+x_2+x_3+x_4+x_5=0,$$
$$-2x_1-x_2+x_4+2x_5=0,$$

or that the totals of observation and expectation have been made to agree, and that the population mean has been estimated from the sample. We solve for x_4 and x_5, express the quantity

$$\sum_{i=1}^{5}\frac{x_i^2}{m_i}$$

in terms of x_1, x_2 and x_3 only, write the expression so obtained as the sum of three linear squares and, after substitution of the new variables z_1, z_2 and z_3, solve for the x's. We thus obtain

$$x_1=\frac{(m_1m_4m_5)^{\frac{1}{2}}}{A^{\frac{1}{2}}}z_1-\frac{m_1m_2^{\frac{1}{2}}(12m_5+6m_4)}{A^{\frac{1}{2}}B^{\frac{1}{2}}}z_2-\frac{m_1m_3(8m_5+3m_4-m_2)}{B^{\frac{1}{2}}C^{\frac{1}{2}}}z_3,$$
$$x_2=\frac{A^{\frac{1}{2}}m_2^{\frac{1}{2}}}{B^{\frac{1}{2}}}z_2-\frac{m_2m_3^{\frac{1}{2}}(6m_5+2m_4+2m_1)}{B^{\frac{1}{2}}C^{\frac{1}{2}}}z_3,$$
$$x_3=\frac{B^{\frac{1}{2}}m_3^{\frac{1}{2}}}{C^{\frac{1}{2}}}z_3,$$

where

$$A=m_4m_5+16m_1m_5+9m_1m_4,$$
$$B=A+m_2(9m_5+4m_4+m_1),$$
$$C=B+m_3(4m_5+m_4+m_2+4m_1).$$

The cosine of the angle between the planes $x_1=0$, $x_2=0$ is

$$r_{12}=-\frac{(12m_5+6m_4+2m_3)(m_1m_2)^{\frac{1}{2}}}{F^{\frac{1}{2}}G^{\frac{1}{2}}},$$

where
$$F = m_5m_4 + 4m_5m_3 + 16m_5m_1 + m_4m_3 + 9m_4m_1 + 4m_3m_1,$$
$$G = m_5m_4 + 4m_5m_3 + 9m_5m_2 + m_4m_3 + 4m_4m_2 + m_3m_2.$$

Again it will be noted that owing to symmetry in the restraints r_{45} may be deduced from r_{12} by the substitution of the appropriate indices.

6. It is seen, as in fact is expected, that the correlation between x_i and x_j for more than one restraint depends on the shape of the population from which the sample has been drawn, i.e. it depends on the relative values of $m_1, m_2, \ldots$, and it will also depend on the type of restraint which is placed on the sample. The behaviour of these correlations is interesting and study of them arose out of a larger investigation into the relations between the signs of the deviations as the number of moment-restraints is increased. Thus it is found that the correlation between x_1 (say) and x_2, negative for one restraint, increases numerically to -1 as the number of moment-restraints increase. On the other hand, the correlation between x_1 and x_3, negative for one restraint, decreases numerically and finally increases to $+1$ with increasing numbers of moment-restraints. Generally we may expect the correlations between two x's with odd subscripts or two x's with the even subscripts to tend to $+1$, and the correlations between an x with an odd subscript and an x with an even subscript to tend to -1.

7. Several tests regarding the signs of the deviations, x, have been proposed recently and the opinion has been expressed that these signs could be regarded as effectively independent for the case where more than one restraint is placed on the material. For such sign tests what is important is the correlation between adjacent deviations and this will, for the case of moment-restraints at any rate, increase numerically to -1; for the case of extreme deviations this increase will be a rapid one. For illustrative purposes I have considered two cases, the first where all the m's are given equal weight, i.e. letting $m_1 = m_2 = m_3 = \ldots = m_k$ and second where $m_1 = m_k < m_2 = m_{k-1} < m_3 = m_{k-2}$* etc. For the first case the correlation between x_1 and x_2 was found for five groups and 1, 2, 3 moment-restraints, seven groups and 1, 2, 3, 4 moment-restraints, nine groups and 1, 2, 3, 4 moment-restraints. The correlation coefficients are given in Table 1.

Table 1. *Correlation between x_1 and x_2; equal weighting of expectations*

No. of groups	5			7				9			
No. of restraints	1	2	3	1	2	3	4	1	2	3	4
Correlation, r_{12}	−0·25	−0·76	−0·92	−0·17	−0·58	−0·87	−0·95	−0·12	−0·43	0·80	−0·92

For the second case the correlations were compared for five groups only and the results are given in the last row of Table 2.

Table 2. *Correlation between x_1 and x_2; expectations weighted equally and unequally*

No. of groups	5		
No. of restraints	1	2	3
Correlation, r_{12}: Equal weights	−0·25	−0·76	−0·92
Unequal weights	−0·12	−0·57	−0·94

The precise numerical value of these correlation coefficients is not important; what is important is the rapidity with which the correlation increases as the number of moment-restraints is increased. This rapid increase would suggest that any test derived on the assumption of the independence of signs of deviations, for extreme observations at any rate, is of doubtful validity.

8. As a check on theory I considered the first 100 samples of the 208 samples, each of 200 observations, used by Neyman & Pearson (1928) as an illustration in their χ^2 paper. The population in this case is a

* Actually the frequencies were obtained by dividing the normal curve between mean $\pm 3\sigma$ into groups with equal base range.

cubic divided into eight groups. I have shown previously (David, 1947) that for the case where one restraint is placed on the material the order of the signs of the deviations can be regarded as random, thus indicating that the effect of the correlation is not felt. We now consider the case when four restraints are placed on the material; the totals of expectation and observation are made to agree and the first three moments are estimated from the data. The correlation between x_1 and x_2 is now equal to $-0{\cdot}840$, and the correlation between x_1 and x_3 is $+0{\cdot}213$. It will be noted that r_{13} is comparable in magnitude with the correlations given in Table 1 for 4 restraints and equal weighting of the expectations. It is clear that both these correlations are too large to be neglected, that between x_1 and x_2 being of sufficient magnitude seriously to invalidate any sign test which is based on a hypothesis of independence or of randomness.

9. It will be realized that the manner in which restraints are imposed on the differences $x_i = n_i - m_i$ is not quite the same in the problem we have been considering as in that arising in tests where a theoretical law is fitted to the data. In the former case the x_i satisfy certain linear relations because we confine attention to the set of samples for which the n_i satisfy certain conditions. In the second case the n_i are unrestricted except for the condition

$$\sum_{i=1}^{k} n_i = N,$$

but in calculating x_i we substitute $\hat{m}_i$ for m_i, where these estimates are so chosen that $n_i - \hat{m}_i$ satisfy similar conditions.

This question has been discussed at some length only by Neyman & Pearson (1928) and is invariably inadequately treated in statistical text-books. I do not propose to discuss the matter fully here but would offer certain remarks which lead me to believe that the substitution of sample estimates, $\hat{m}$, for the population values, m, in (1) will not seriously invalidate the calculated correlations.

We may consider for simplicity a population of four groups. As before, we shall have

$$x_i = n_i - m_i,$$

and we shall write

$$X_i = \hat{m}_i - m_i = f_i(\phi),$$

where ϕ is the estimate of the population parameter which is calculated from the observations. Because ϕ is an estimate it will vary from sample to sample and the estimated values X_1, X_2, X_3, X_4 will lie on a curve (termed by Neyman & Pearson the population locus), which will depend on ϕ alone. It is clear therefore that for any given set of observations x_1, x_2, x_3, x_4 the estimated population point will depend on the method of fitting used, or, perhaps more precisely, on the method of estimation of ϕ.

If the structure of the restraint placed on the observations for the purpose of estimating ϕ is such as to lead to a minimum

$$\chi^2 = \sum_{i=1}^{4} \frac{(n_i - \hat{m}_i)^2}{\hat{m}_i},$$

then the correlations, obtained under the assumption that (1) is true when $\hat{m}_i$ is substituted for m_i, will be approximately correct. In general the method of moments will not lead to this minimum χ^2 exactly, but the fitted expectations will not usually be very different from the minimum values, and the error made in assuming (1) is true will not be large.

REFERENCES

David, F. N. (1947). *Biometrika*, **34**, 299.
Neyman, J. & Pearson, E. S. (1928). *Biometrika*, **20A**, 274.

Note on 'Proofs of the distribution law of the second order moment statistics'

By JOHN WISHART

Since the paper under the above title was published (1948, *Biometrika*, **35**, 55), my attention has been called to a paper by E. Sverdrup, 'Derivation of the Wishart Distribution of the Second Order Sample Moments by straightforward Integration of a Multiple Integral' (1947, *Skand. AktuarTidskr.* **30**, 151). A copy has also been received of a paper by R. D. Narain, 'A New Approach to Sampling Distributions of the Multivariate Normal Theory', in which the same distribution is derived. This paper will shortly be published in India.

REVIEW

Probit Analysis. By D. J. Finney, M.A. Cambridge University Press. 1947. Price: 18*s*.

The publication of books relating to specialized statistical techniques is a welcome feature of recent years. In such books it is possible to give an account sufficiently detailed to provide a practical guide to the user of these techniques in most of the problems to be encountered, as opposed to the necessarily more restricted treatment of any particular topic in a general text-book. Mr Finney's volume on probit analysis, with special emphasis on applications in biological research, is an excellent example of this specialized type of book, not only taking full advantage of the possibility of exhaustive treatment, but also presenting the arguments in such a way as to ease the reader's task in mastering the methods described.

Starting from a simple description of the types of problem to be treated, techniques covering almost every eventuality in the field commonly described as 'dosage-mortality' are developed and their practical application illustrated by means of numerical examples. The standard methods of probit analysis are carefully described and very useful sets of tables (Tables I–IV) are provided. Tables III and IV, in particular, should considerably facilitate the computation of working probits. Table III gives both maximum and minimum working probits for expected probits at intervals of 0·1; Table IV enables the working probit to be obtained directly from the provisional probit and the percentage 'kill'. Table II, apart from giving values of Q/Z, gives values of the weighting coefficient allowing for the effect of natural mortality. Chapter 6 is devoted to the consideration of methods of allowing for natural mortality and should prove most useful. While the methods described are not new, it is the first time that they have appeared in a systematic treatise in such a way as to bring them to general notice. Similar remarks might be made about the other special features of the book—the treatment of factorial experiments, of the joint action of different poisons and of quantitative responses, for example. Most of the methods described have already been published in individual papers, but Mr Finney has performed an important task in bringing them together in one consolidated account. The method of development is not, of course, theoretically complete or rigorous, but it is logically clear, and should provide an invaluable aid to experimentalists in their appreciation of the statistical tests described.

It may seem ungenerous to extract for special criticism the very few passages in this useful book which seem to be ambiguous or capable of improvement. Such criticism will be made, however, as a modest attempt to help readers of the work to avoid certain difficulties and confusion which they might possibly experience. It appears to the reviewer that some difficulties may result from the author's desire to simplify the presentation of the underlying theory. While it must be admitted that some simplification is necessary, this should not be such as to allow of false impressions of the theoretical situation. As it appears that considerable knowledge of general statistical theory is expected of the reader, there would seem to be no need to avoid the use of simple technicalities in clarifying the exposition.

We may instance the arguments for introducing the heterogeneity factor (p. 33), the working probit (p. 48) and the correction for natural mortality (p. 88). In the first of these cases it should be noted that the interval of estimation will be on the average too long if there is no heterogeneity; while if heterogeneity is present the occasional use of normal factors will be incorrect. On p. 48, the implication that it is merely the asymmetry of the distribution of p which gives rise to the need for working probits gives a rather incomplete notion of the theory. Finally, the arguments on corrections for natural mortality could be much simplified if they were based directly on equation (6·2).

The use of both LD 50 and ED 50 to symbolize the same population parameter, LD 50 being restricted to cases where the response is death, may be somewhat confusing. We note, indeed, that the author himself uses the symbol ED 50 on p. 124, when considering 'kills'. It may save some confusion to point out that the quantity x_1x_2, defined as x_3 on p. 117, is afterwards represented as x_{12} (there is also a misplaced bracket in the first equation on this page); there is also a misprint on p. 89, where the reference to equation (3·5) should be to equation (3·4).

The comparatively minor importance of the above criticisms must be emphasized. The book as a whole is a valuable addition to the statistician's library and is indispensable for persons concerned with 'dosage-mortality' problems, in whatever form they arise. Apart from its other merits, the very full bibliography makes the book invaluable as a reference index in its own field. May other works of this specialized nature maintain the standard set by Mr Finney! N. L. J.

CORRIGENDA

1. Moments of the mean deviation from the mean in samples from a normal population. (See the reference in paper by H. J. Godwin, p. 308 above.) The following corrections should be made:

(1·1) R. C. Geary (1936), *Biometrika*, **28**:

p. 300 In equation (22) the coefficient of n^{-4} in the expansion for m_4'' should read

$$(51-352a^2+\tfrac{2136}{5}a^4), \quad \textit{not} \quad (51-352a^2+427a^4).$$

p. 301 In equation (24) the coefficient of n^{-4} in the expression for $m_4''-m_4'$ should read

$$0{\cdot}11463500, \quad \textit{not} \quad 0{\cdot}033\dot{5}7805.$$

(1.2) E. S. Pearson (1945), *Biometrika*, **33**, p. 252:

In equation (4) the coefficient of ν^{-4} in the expansion for λ_4 should read $-0{\cdot}038946$, *not* $-0{\cdot}120003$. This correction makes alterations in the values for λ_4 and β_2 given in the table of moments of the mean deviation on p. 253, which should read:

Sample size n	λ_4	β_2
4	0·001 963	3·252
5	·000 9912	3·197
6	·000 5672	3·161
8	·000 2356	3·118
10	·000 1195	3·093
12	·000 06868	3·076
15	·000 03493	3·061
20	·000 01464	3·045

2. *Biometrika*, **35**, 181 (1948):

In the list of references at the end of F. Yates's paper, that to E. S. Pearson (1923) should read *Biometrika*, **14**, 261, *not Biometrika*, **7**, 248.